Check for updates

Review of few-shot learning application in CSI human sensing

Zhengjie Wang[1] · Jianhang Li[1] · Wenchao Wang[1] · Zhaolei Dong[1] · Qingwei Zhang[1] · Yinjing Guo[1]

Abstract

Wi-Fi sensing has garnered increasing interest for its significant advantages, primarily leveraging Wi-Fi signal fluctuations induced by human activities and advanced neural network algorithms. However, its application faces challenges due to limited generalizability, necessitating frequent data recollection and neural network retraining for adaptation to new environments. To address these limitations, some researchers introduced few-shot learning into Wi-Fi sensing applications because it offers a promising solution with its ability to achieve remarkable performance in novel scenarios using minimal training samples. Despite its potential, a comprehensive review of its applications within this domain remains absent. This study endeavors to fill this gap by exploring prominent Wi-Fi sensing applications that incorporate few-shot learning, aiming to delineate their key features. We categorize few-shot learning approaches into three distinct methodologies: transfer learning, metric learning, and meta-learning, based on their neural network training strategies. Through this classification, we examine representative systems from an application perspective and elucidate the principles of few-shot learning implementation. These systems are evaluated in terms of learning methodology, data modality, and recognition accuracy. Finally, this paper highlights the challenges and future directions for few-shot learning in Channel State Information (CSI) based human sensing, providing a valuable resource for researchers in the field of Wi-Fi human sensing leveraging few-shot learning.

1 Introduction

Wireless sensing technology stands out as a popular research direction due to its convenience compared to wearable sensors and reduced privacy concerns compared to visual approaches. Wireless sensing devices mainly include millimeter-wave radar and Wi-Fi devices. Wi-Fi signal has many advantages since it can utilize existing commercial Wi-Fi

---

✉ Yinjing Guo
  gyjlwh004@126.com

1   College of Electronic and Information Engineering, Shandong University of Science and Technology, Qianwangang Road, Qingdao 266590, Shandong, China

equipment. It achieves human sensing by processing real-time channel state information (CSI) related to the environment, minimizing privacy leakage in the detection process. The key concept behind Wi-Fi sensing is the user's movement (i.e., people or other objects), as shown in Fig. 1. Different motion patterns exhibit distinct characteristics that can be utilized for various applications in detection estimation, activity recognition, and fall detection (Ma et al. 2019; Wang et al. 2021a).

In recent years, with the rapid development of deep learning (Khan et al. 2020), more and more research has been carried out on the feature extraction and classification of CSI using deep learning methods (Zhang et al. 2022a; Abdelnasser et al. 2015; Li et al. 2016; Venkatnarayan et al. 2018). However, existing methods heavily rely on the data collected from the source environment, and the lack of sufficient training data limits the model's generalization ability. When a well-trained model encounters the data from different environments, its accuracy may decrease significantly. Additionally, previous work has indicated that the mapping between human activity and the resulting signal variations is not bidirectional (Gao et al. 2021; Niu et al. 2022; Chen et al. 2023). When each activity is performed at different positions relative to the Wi-Fi transceiver, it may lead to different signal variations.

We refer to factors causing signal fluctuations other than human movement as "domain". Whenever there is a change in any domain, it causes fluctuations in the Wi-Fi signals. Therefore, additional efforts are required to collect and label data when a new domain emerges. Failure to update the data from the new domain will result in unreliable sensing performance. However, it is impractical to collect data from an infinite number of domains. As a result, when using a model trained in the source domain to recognize action categories in the target domain, the recognition accuracy may drop significantly.

Cross-domain recognition is a challenging problem in CSI-based human activity recognition (HAR) due to significant differences in data collected from different environments and scenarios. Traditional deep learning algorithms rely on large amounts of context-specific labeled data, limiting their ability to perform well in cross-domain settings. Therefore, CSI-based human sensing cross-domain recognition becomes a current challenge.

To address this challenge, many scholars have conducted extensive research on the issue of cross-domain adaptation, and two primary approaches have emerged. On the one
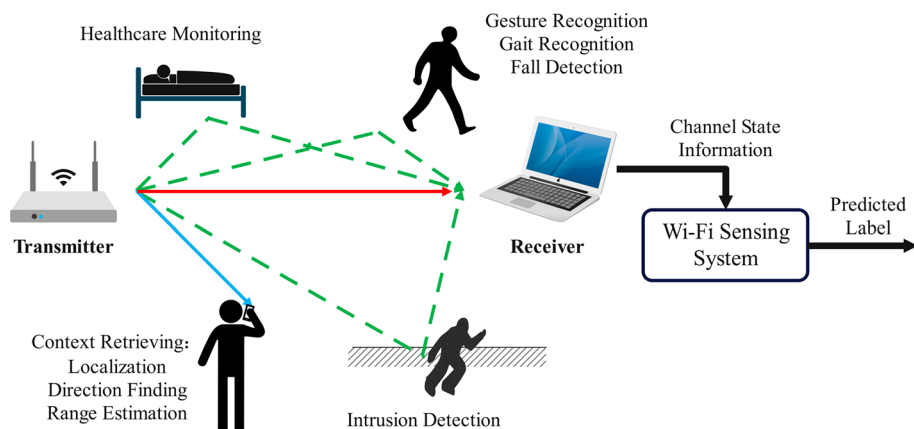
Fig. 1 An overview of applications based on Wi-Fi sensing

hand, features independent of the domain are extracted. For example, Zhang et al. (2022b) propose to extract the power distribution of different gesture speeds from the Doppler spectrum and use the time learning model to leverage the extracted features and achieve domain-independent gesture recognition. However, a lot of data pre-processing, complex feature extraction, and training data are required. At the same time, the experimental results show that the algorithm's accuracy decreases obviously when the number of samples is reduced. On the other hand, domain adaptation methods are explored to deal with such problems. For example, Virmani and Shahzad (2017) proposed a transformation method to generate virtual samples of the target domain automatically, and the recognition model is trained using virtual samples under all possible domain configurations. However, there are still many important situations that cannot be adequately taken into account.

In addition to the problem of decreasing recognition rate after cross-domain, which requires adjusting the large amount of data collected by the new domain, there is also a common disadvantage: the inability to recognize new categories of activities. When further action is introduced, the entire model must be retrained using all the training data. Collecting data and retraining models take much time, and this drawback greatly hinders their use in the real world, as predefined collection conditions cannot meet the growing number of requirements.

Reducing data collection in new perceptual scenarios while maintaining a high recognition rate has become a research topic in current cross-domain and new activity recognition. Fortunately, few-shot learning is what we need exactly.

Few-shot learning refers to a learning technique that rapidly adapts to unseen tasks with only a few available samples. In other words, designers do not need to worry too much about the quantity of data. This method is inspired by human learning. For example, toddlers can recognize new object categories with just a few examples. Typically, when learning new tasks, individuals leverage previous knowledge and experiences, adapt to the new tasks based on the provided context, and induce abstract knowledge about how to learn, enabling them to learn relevant new tasks and effective adaptation rapidly.

Few-shot learning(Fe-Fei and Fergus 2003) has been proposed to learn from a small number of labeled samples and has shown significant improvements in interpreting natural images, including image classification(Liu et al. 2022), object detection(Antonelli et al. 2022), and activity recognition(Wang et al. 2023). Unlike traditional deep learning approaches, few-shot learning can train high-performance classifiers using only one or a few labeled data points. The key to few-shot learning lies in comparing the similarity of data across different domains. Unlike methods that rely on a large number of samples from new classes, few-shot learning techniques leverage prior knowledge gained from previous experiences to facilitate rapid learning of new tasks. Inspired by its success in computer vision, researchers have extensively explored effective few-shot learning methods for human sensing based on CSI. Few-shot learning methods can help reduce the collection of target domain data and increase the generalization for different commercial scenarios.

Numerous studies have provided detailed explanations of few-shot learning(Wang et al. 2020) and presented its application in various scenarios, such as image classification(Liu et al. 2022) and object detection(Huang et al. 2023). Additionally, several works have applied few-shot learning to sensor-based human activity recognition(Gupta et al. 2022; Khan and Ghani 2021) and utilized meta-learning(Halperin et al. 2011; Xue et al. 2023) for optimizing signal processing. Previous research has already demonstrated the capabilities of Wi-Fi-based sensing systems across various applications(Chen et al. 2023). While related works have summarized cross-domain research on Wi-Fi sensing(Koch et al. 2015), their investigation was

primarily concentrated on metric learning in few-shot learning environments. However, there remains a limited in-depth exploration of the application of few-shot learning in CSI human sensing. To address the current research gap in cross-domain human sensing of CSI using few-shot learning, this paper reviews the latest research progress of few-shot learning in CSI human sensing and provides an outlook on future research directions. Our objective is to provide readers with a more comprehensive understanding of the development of few-shot learning in cross-domain human sensing of CSI, enabling better application in real-world scenarios.

The main contributions of this work are as follows:

(1) To the best of our knowledge, this paper is the first comprehensive review of human behavior sensing based on CSI and few-shot learning. It emphasizes the crucial technology of how to utilize few-shot learning techniques to harness the distinctive characteristics of CSI data effectively.
(2) The paper introduces the typical few-shot learning theory used in CSI sensing. Then, typical human sensing application cases are presented, including gesture, activity, localization, and crowd counting. This study delves into the critical few-shot learning models underpinning these applications, offering a detailed examination of their methodologies and effectiveness.
(3) The paper identifies and outlines the crucial challenges encountered when applying few-shot learning to CSI-based human sensing. Furthermore, it presents a discussion on prospective research directions, aiming to illuminate pathways for future investigations and advancements in this field.

For a better reading, we summarize the following chapters with flowcharts Fig. 2. Section 2 provides CSI-related information and preprocessing. Section 3 explicitly introduces few-shot learning and related classical networks. Section 4 presents the typical application of few-shot learning in CSI human sensing. Section 5 presents the shortcomings and challenges of the current development. Section 6 concludes with a final remark.
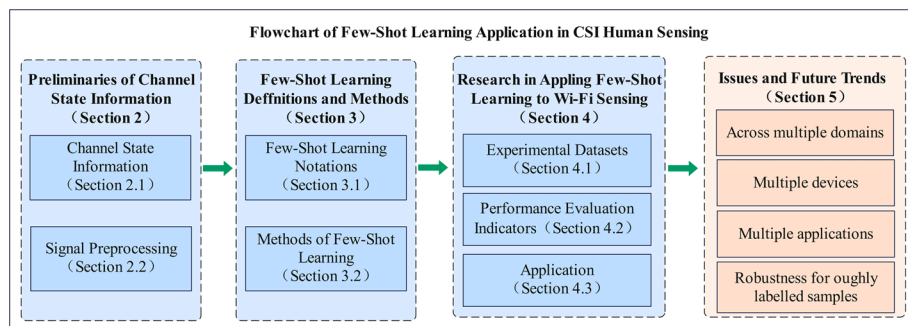
Fig. 2 Flowchart of this survey

2 Preliminaries of channel state information

This section provides a brief introduction and summary of the CSI-related concepts, data collection devices, and preprocessing. Typically, CSI collection and processing are performed by devices equipped with network interface cards (NICs). Subsequently, it is necessary to extract the selected fundamental signals, such as amplitude or phase, from the collected information. In the next step, the extracted signals are fed into a signal preprocessing module to remove noise from the signals, obtaining more accurate CSI data.

2.1 Channel state information

CSI exhibits the propagation characteristics of a wireless signal as it traverses multiple paths from the transmitter to the receiver at a specific data rate and carrier frequencies. The time series measured by CSI captures how wireless signals propagate through surrounding objects and people in time, frequency, and space so that they can be used to keep stable communication.

The characteristics of the physical layer measurement sub-carrier channel are obtained by extracting the channel state information in the Wi-Fi signal. Then, the complex multipath effects caused by human motion are revealed to realize the detection and sensing of the human body. At present, most commercial off-the-shelf (COTS) Wi-Fi routers are designed for multiple-input multiple-output (MIMO), multi-antenna communication and generally use orthogonal frequency division multiplexing (OFDM) technology, supporting IEEE 802.11n/ac/axe standards. The data rate is increased by transmitting many narrowband carriers at different frequencies simultaneously, so the CSI includes amplitude attenuation and phase offset of multiple paths in each subcarrier.

The CSI can describe the time delay, attenuation, and phase shift during signal propagation. It can be defined in the frequency domain by the following formula.

$$Y = HX + G \tag{1}$$

where Y and X are the receive and transmit signal vectors, respectively, G is the additive Gaussian white noise vector, and H is the complex matrix representing CSI.

Considering a Wi-Fi system operating under IEEE 802.11n specification, and with M transmitting antennas and N receiving antennas, the signal that contains the estimated CSI of each data stream can be mathematically expressed as
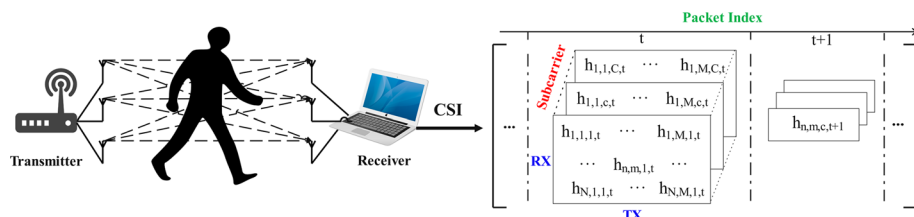
$$H = \begin{pmatrix} h_{1,1} & h_{1,2} & \dots & h_{1,M} \\ h_{2,1} & h_{2,2} & \dots & h_{2,M} \\ \vdots & \vdots & \dots & \vdots \\ h_{N,1} & h_{N,2} & \dots & h_{N,M} \end{pmatrix} \tag{2}$$

For each pair of receive and transmit antennas can be written as

$$h = [h_1, h_2, \dots, h_C] \tag{3}$$

where C is the number of subcarriers. Meanwhile, h_C can be expressed as

$$h_C = \|h_C\| e^{j sin \angle h_C} \tag{4}$$

Fig. 3 N×M MIMO-OFDM Wi-Fi human sensing data acquisition schematic diagram

The CSI signal can be represented as a complex 4D tensor, $H \in \mathbb{C}^{M \times N \times C \times T}$, where M is the number of transmit antennas, N is the number of receive antennas, C is the number of sub-carriers, and T is the sampling time.

We can consider a typical Wi-Fi human sensing scenario, where the router with M antennas serves as the transmitter and a laptop equipped with N antennas as the receiver, as shown in Fig. 3. Where m is the mth transmitting antenna, n is the nth receiving antenna, t is a certain time, and c is the c th subcarrier.

The number of subcarriers is determined by bandwidth and tools. The most commonly used CSI tools are the Intel 5300 NIC, Atheros CSI, and Nexmon CSI. The Intel 5300 NIC was the first and most widely used tool for collecting CSI. It can capture 30 subcarriers for each pair of antennas operating at 20MHz bandwidth. The Atheros CSI tool increases the CSI to 56 subcarriers at 20MHz and 114 subcarriers at 40MHz to improve the resolution of the CSI data.

For the first time, the Nexmon CSI Tool enables CSI capture on portable devices such as smartphones and Raspberry Pi. It can capture 256 subcarriers at 80MHz. However, it has protection and empty subcarriers (Gringoli et al. 2019) that must be removed before signal processing. In addition to the above three, with the development of CSI sensing, the tools and supporting devices to support the capture and collection of CSI are increasing. Table 1 briefly introduces the information on relevant collection tools.

Table 1 Collection device information

Tools	Support hardware	Max subcarriers
Linux 802.11n CSI Tool (Yang et al. 2018)	IWL5300	30
ESP32 CSI Tool (Hernandez and Bulut 2020)	ESP32	64
Atheros CSI Tool (Yang et al. 2018)	AR9580, AR9590	114
Nexmon CSI Tool (Gringoli et al. 2019)	BCM4330, BCM4339, BCM4358, BCM4343	Up to 4096
PicoScenes (Jiang et al. 2022)	AX200, AX210, USRP, QCA9300, IWL5300	Up to 7968
AX-CSI (Gringoli et al. 2022)	BCM43684	Up to 32768

Table 2 Signal preprocessing techniques for CSI sensing

Method	Examples
Noise reduction	Filters: (Ding et al. 2022; Ma et al. 2020; Zhang et al. 2022f; Zhou et al. 2022) (Xiao et al. 2021; Ding et al. 2021; Hu et al. 2021; Yin et al. 2022) (Zheng et al. 2023b) Phase Unwrapping: (Zhang et al. 2022f; Zhou et al. 2022) (Hu et al. 2021) Wavelet Transform: (Wang et al. 2024, 2021b)
Data adaptation	PCA: (Zhou et al. 2022; Xiao et al. 2021; Shi et al. 2022; Gu et al. 2021) (Zheng et al. 2023b), SVD: (Bahadori et al. 2022)
Signal transform	FFT: (Shi et al. 2022) STFT: (Zhou et al. 2022; Xiao et al. 2021) DWT: (Gu et al. 2021)

2.2 Signal preprocessing

In general, accurate human identification requires the collection of precise data describing human behavior. The CSI data for wireless signals include proper signals, unordered noise, outliers due to complex environments, signal interference, and moving people. Therefore, data preprocessing methods are crucial, as shown in Table 2. The relevant data preprocessing methods are briefly introduced, including noise reduction, data adaptation, and signal transformation.

2.2.1 Noise reduction

The noise source is very complex, including hardware factors such as central frequency offset (CFO) and sampling frequency offset (SFO) errors, as well as environmental factors such as signal shadows and multipath fading. These factors cause the signal to travel through multiple non-line-of-sight paths to the receiving antenna, leading to destructive interference. Noise reduction is typically performed independently for each subcarrier.

It is common to filter the signal and process the threshold using different filter algorithms. These include frequency response filters, Butterworth filters, moving average filters, bandpass filters, etc. In addition to the classic filters mentioned above, conjugate multiplication is also used to filter out irrelevant noise and retain the necessary information. When different antennas on the Wi-Fi card share the same oscillator, the time-varying random phase offset is the same, and one antenna is selected as the reference antenna to calculate the conjugate multiplication. In addition, much work is done using mathematical operations to reduce noise while addressing offset noise and multipath interference, such as phase unwrapping and ratio calculation. Wang et al. (2018a) first used phase unwrapping to derive the adjusted phase of each subcarrier of CSI from realizing a Wi-Fi-based material detection system. FingerDraw (Wu et al. 2020) proposes a CSI ratio operation. By calculating the quotient of two CSI signals from different antennas of the same receiver, the random phase offset of the antennas of the same receiver is eliminated, and the signal-to-noise ratio(SNR) is effectively maximized.

2.2.2 Data adaptation

Each collected CSI sample contains a complex subcarrier vector. Tan et al. (2022) have demonstrated that some subcarriers have similar properties and contain redundant information, while others are subject to large amounts of noise. Abdelwahed (Khamis et al. 2020) selects the subcarrier by considering the statistical characteristics of each subcarrier over a predefined time frame. In addition, many studies use dimensionality reduction algorithms to eliminate these redundant subcarriers.

Traditional compression algorithms include principal component analysis (PCA) using linear transformations and singular value decomposition (SVD). PCA is widely used to reduce the dimensionality of data while preserving most of the information about the selected central component, a linearly uncorrelated and ordered set of variables sorted by the proportion of total information each variable contains. Most existing work chooses to retain the most information about the first principal component. Similarly, SVD (Bahadori et al. 2022) for data dimensionality reduction.

2.2.3 Signal transform

Traditionally, amplitude and phase are used to complete the subsequent activity identification tasks, and the frequency component is ignored. The frequency component is a good characterization because different movements have dominant frequencies. However, the original CSI measurements only show the amplitude and phase changes over time, not the frequency components.

Fast Fourier transformation (FFT) is the most common method to convert CSI measurements from the time domain to the corresponding frequency domain. The FFT can also be used to obtain the power spectral density (PSD), which has been used to estimate respiration/heart rate (Wang et al. 2024). However, FFT needs more information in the time domain. Short-time Fourier transform (STFT) and discrete wavelet transform (DWT) can capture time and frequency domain features. The STFT slides a window over the time series measured by the CSI. At each sliding step, it applies the FFT to the CSI value of the window covering. Thus, the window size determines the trade-off between STFT frequency and temporal resolution. The larger the window, the higher the frequency resolution of the STFT and the lower the time resolution, and vice versa. DWT is based on multi-resolution analysis, providing high time resolution for high-frequency motion and high-frequency resolution for low-frequency signal.

CSI-based human sensing with few-shot learning has the same network architecture as computer vision with few-shot learning. The main difference between the two applications is the input data: wireless signals vs. 2D images. Because CSI data has distinct characteristics compared with traditional computer vision data, the data size of different dimensions is different due to the influence of acquisition equipment. Therefore, for the convenience of neural network input, several approaches were selected for processing the CSI data. For example, (Wang et al. 2022a; Zhang et al. 2022c; Ding et al. 2022; Huang et al. 2022; Bahadori et al. 2022; Ding et al. 2021; Yang et al. 2019; Ma et al. 2020; Wang et al. 2022b; Hou et al. 2022; Wang et al. 2021b; Gu et al. 2021; Zhang et al. 2022d; Gao et al. 2023; Zhang et al. 2022e; Wang et al. 2024; Wei et al. 2023; Zhang et al. 2023a; Hu et al. 2021) directly input the preprocessed signals, (Shi et al. 2022; Zhou et al. 2022; Xiao et al. 2021) transform them into spectrograms for input, (Hou et al. 2023; Zheng et al. 2023b) segment

the data into specified lengths and (Zhang et al. 2022f; Chen and Chang 2022) reshape data to convert it into one-dimensional data.

3 Few-shot learning definitions and methods

For traditional deep learning-based CSI human sensing, WiGRUNT (Gu et al. 2022) extracted a subset of samples from the WiDar3.0 dataset as a dataset and conducted experiments with different locations, environments, and orientations. Accuracy declined compared to the in-domain experiment and further declined as the number of participants and active gestures increased. At the same time, WiGr (Zhang et al. 2022f) demonstrated that accuracy drops below 20% when a model trained on the current location is applied to test data from a new location.

At the same time, signal preprocessing also does not solve the cross-domain problem because the processed signal features are still domain-dependent. A large amount of data must still be collected from the test domain to train the network to maintain accuracy. Given the challenges observed in traditional deep learning-based CSI human sensing, it is imperative to explore innovative approaches. Few-shot learning can effectively utilize prior knowledge and adapt to new tasks with minimal training data, providing an interesting prospect for solving these constraints. By integrating a small number of learning techniques into the CSI human sensing domain, performance can be improved, especially in scenarios with sparse training data. This shift sets the stage for further exploration of the potential applications of small amounts of learning in the context of CSI human sensing.

3.1 Few-shot learning notations

Few-shot learning is the process of training a model with very little training data. The expectation is to learn a priori knowledge from a large number of basic training tasks and to transfer the learned knowledge into a new class consisting of a small number of labeled samples. When the number of training examples is minimal, this method can use previously acquired knowledge to improve performance on new tasks (Xie et al. 2020).

Given a task set T, a few-shot task set $D_{train} = \{T_1, \cdots, T_i\}$ is extracted from the task set T for training. The tasks are independent and have $T \sim p(T)$. There is a task distribution in which each sample T_i is a specific supervised few-shot task. T_i contains two collections: the support set and the query set. The sample labels in the query set are consistent with those in the support set. During training, the samples from the support set are used to minimize the model's classification error on the query set.

In the testing stage, another test set $D_{test} = \{T_1, \cdots, T_i\}$ is sampled from the task set to verify the model's performance on few-shot tasks and to calculate the model's accuracy. The category of few-shot functions in the test set completely differs from that in the training set. Few-shot learning aims to find a model to minimize the expected risk of all few-shot tasks. Assuming that the parameters of the model are θ, then the objective function of its learning is:

$$\min_{\theta} E_{T \sim p(T)} L(T;\theta) \tag{5}$$

Compared to traditional supervised learning methods, few-shot learning has few labeled samples to obtain extensive prior knowledge. Therefore, knowledge transfer and internal relationship learning for samples in the same class become crucial in few-shot learning. Vinyals et al. (2016) propose an episodic training method. In the training process, we extract N categories from the training set, construct a $N \times K$ support set for each category K sample from these data, and randomly select several samples from the remaining data of that category to build a query set for few-shot learning. We can iterate over multiple scenarios to achieve convergence by constructing a query set that assists in few-shot learning. This query set and the support set form a comprehensive set. Few-shot learning is often considered an N-way K-shot problem. N is the number of categories in each task support set, and K is the number of samples in each category.

The core idea of few-shot learning is to expect the model to generalize experience to new task scenarios, just as humans can use experience to learn new knowledge quickly. Specifically, the prior knowledge obtained from the auxiliary dataset, which can exist in various forms (such as parameter initialization, pre-extracted features, etc.), assists the current learning task by designing appropriate learning strategies.

3.2 Methods of few-shot learning

So far, there is no unified and comprehensive standard for classifying few-shot learning methods. Different tasks are classified based on various technical approaches. For instance, Duan et al. (2021) categorized them into three types based on prior knowledge: model-based, data-based, and algorithm-based. The modeling principles of few-shot learning (Lu et al. 2020) can be divided into two categories: model generation methods and model discrimination methods.

In contrast to the above classification, this article does not provide a review of the latest research on few-shot learning but only summarizes the work related to CSI human sensing. Similar to the CSI human sensing based on traditional deep learning, the CSI human sensing based on few-shot learning also consists of data processing and network models. Data processing is similar to the traditional deep learning method, hoping to reduce the influence of noise and so on, but for the network model, few-shot learning hopes to learn feature selection and processing ability from more limited data.

Research on CSI human sensing that employs few-shot learning has primarily advanced along two distinct categories. One focuses on refining network parameters without altering the network architecture, and another seeks enhancement through architectural modifications of the network itself. In this paper, we classify networks from two aspects: parameter adjustment and network structure. For parameterized size, there are transfer learning and meta-learning. For architectural design, metrics learning is most commonly used.

Transfer learning focuses on how to leverage knowledge learned from existing CSI activity datasets to help solve new tasks, while meta-learning aims to understand how to adapt quickly to new tasks by quickly learning from a small number of CSI activity samples collected in a new domain. Metric learning focuses on learning appropriate metric functions to quickly measure the dissimilarity of different activities in new domains in CSI datasets; we summarize these three methods in Table 3, listing their advantages and disadvantages.

To facilitate a better understanding of the following figure, we provide the definitions of some symbols: S represents a support set, Q represents a query set, X represents a batch of

Table 3 Pros and cons of different learning methods

Methods	Overview	Pros	Cons
Transfer learning	Using trained parameters	Simple application and high efficiency	The source domain dataset is required to be highly correlated with the target domain dataset
Meta-learning	Learning at the task level, accumulating knowledge unrelated to the task to enhance learning	Strong cross-domain generalization and fast adaptation	Requires support for secondary datasets
Metric learning	The samples are embedded into the optimal metric space, and the comparison is performed based on the similarity	In the absence of secondary adaptation, it is generally applicable to multiple tasks of different categories after training	Performance degradation when dealing with new tasks with large distribution differences from the training set

instances, $\hat{y}$ represents a prediction category, f_θ represents feature extraction parameters, g^θ represents the classifier, and M represents the metric function.
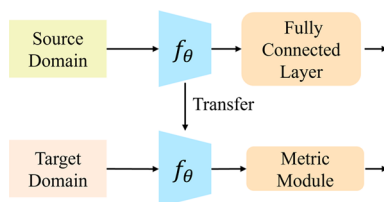
3.2.1 Transfer learning

Transfer learning is an important technique to improve the learning of the target domain by transferring knowledge from the relevant source domain. Most of the CSI human sensing tasks (Yin et al. 2022; Hou et al. 2022, 2023; Wang et al. 2024; Xiao et al. 2021; Gu et al. 2023) based on transfer learning are accomplished through model-based transfer learning, with a few utilizing specifically designed strategies (Wei et al. 2023; Wang et al. 2022a).

For model-based transfer learning, there are typically two strategies: fixing the parameters of the pre-trained model as feature extractors or fine-tuning these model parameters. FewSense(Yin et al. 2022) discusses both of these strategies. The architecture of such methods, as illustrated in Fig. 4, involves using models obtained traditionally, followed by the use of a metric function to replace the original fully connected layer. Various studies primarily focus on improving the training process and metric models. For example, AutoFi (Yang et al. 2023) introduces self-supervised learning to extract deeper features, and Wang et al. (2024) focus on metric learning for local and global features.

The second approach involves aligning feature representations between domains through the design of specific strategies to share knowledge in the form of shared feature representations, as shown in Fig. 5. By focusing the model on learning shared feature representations in the source domain, the need for labeled training samples in the target domain can be reduced. Wei et al. (2023); Wang et al. (2022a) introduce maximum mean discrepancy (MMD) (Tzeng et al. 2014) for feature alignment across different domains.

Wi-Fi signal characteristics vary in physical spaces and environments, and transfer learning allows models to adapt to these changes without the need to collect and train large amounts of new data from scratch. While transfer learning in CSI human sensing can facilitate models to achieve better recognition results in few-learning, current methods may need to be optimized. Due to the limited training samples, poor recognition performance

Fig. 4 Overview of the fine-tuning based on transfer learning

**Fig. 5** Overview of the transfer knowledge based on transfer learning
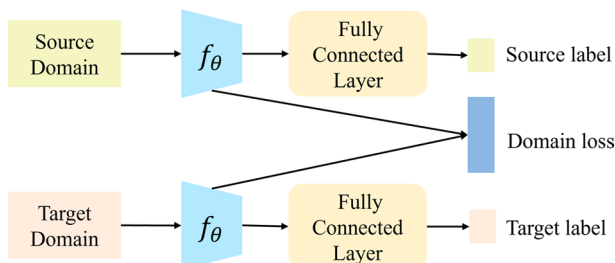
may occur in unknown classes. If the feature extraction model is not fine-tuned (Zhang et al. 2022f), the recognition rate will decrease significantly.

3.2.2 Metric learning

Utilizing metric learning for CSI human sensing has been achieved by methods referenced in (Zhou et al. 2022; Yang et al. 2019; Ma et al. 2020; Shi et al. 2022; Ding et al. 2021; Zhang et al. 2022f; Yang et al. 2023; Ding et al. 2022; Bahadori et al. 2022; Ding et al. 2021; Wang et al. 2022b; Zhang et al. 2022c, 2023a; Hu et al. 2021). Metric learning approaches provide a method for learning an embedding space for each sample, enabling representatives of the same class to be close together. We summarize the existing CSI human sensing work, with a framework outlined in Fig. 6, where neural networks map them to a high-dimensional space and then use a metric function for classification.

For common metric models, they can be classified into two categories: nonparametric-based and parametric-based methods. Nonparametric-based methods compute distance using fixed calculation methods, such as cosine distance and Euclidean distance. On the other hand, parametric-based methods utilize deep learning methods and convolutional layers to measure distances. The commonly used metric models are shown in Fig. 7.

For a better understanding of metric learning, we further explain it in detail. In addition to the different metric approaches mentioned above, many few-shot metric learning methods compare query samples with class representations (e.g., prototypes and sub-spaces) rather than individual samples. This can be categorized into three modes: learning feature embeddings, learning class representations, and learning metrics.
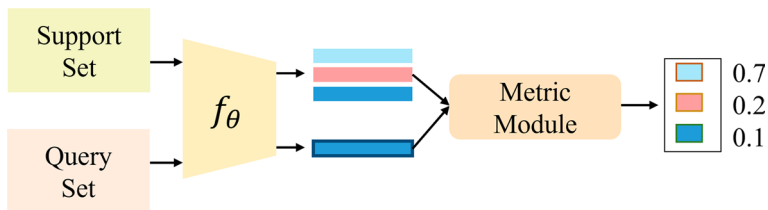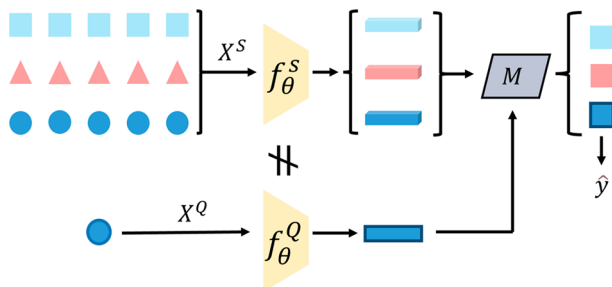

**Fig. 6** Illustration of metric-learning-based few-shot learning methods
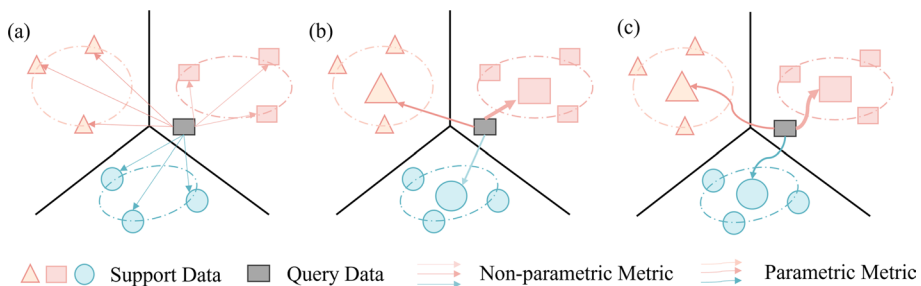
Fig. 7 Illustration for metrics in embedding space. From left to right: **a** Matching Net, **b** Prototypical Net, **c** Relation Net

**Fig. 8** Overview of the matching network(3-way 5-shot task for example)

Methods for learning feature embeddings are considered efficient at extracting discriminative features and generalizing well to new classes. The Siamese neural networks used by (Zhou et al. 2022; Yang et al. 2019), as well as the matching networks utilized by (Ding et al. 2021; Shi et al. 2022), are representative network architectures for such methods. Snell et al. (2017) first used the Siamese neural network as a feature extractor for few-shot learning. The main idea is to use the Siamese network to extract features and calculate the part L1 distance between two samples. If the elements belong to different classes, they will be very far apart. Otherwise, they belong to the same category. Vinyals et al. (2016) proposed a matching network that used metric learning for few-shot image classification and added attention and external memory mechanisms. The network considers images and other images within the image set and classifies the extracted features using cosine similarity (as shown in Fig. 7.a). The proposed matching network structure is shown in Fig. 8.

For learning class representation, the prototype network(Snell et al. 2017) is a classic model. The introduction of the prototype network to CSI human sensing was first done by (Zhang et al. 2022f). Building upon the prototype network, Wang et al. (2022b) introduces open-set recognition to discriminate against unseen classes. The prototype network assumes that each category has its prototype representation in the embedding space. It maps support data to the embedding space and computes the average embedding features of the support data for each category to derive the prototype for each type. In the embedding space, a fixed distance function (such as Euclidean distance) is used to calculate the distance between the query sample and the class prototype (as shown in Fig. 7.b). This distance serves as a measure of similarity between the query sample and the class prototype. The proposed prototype network structure is illustrated in Fig. 9.
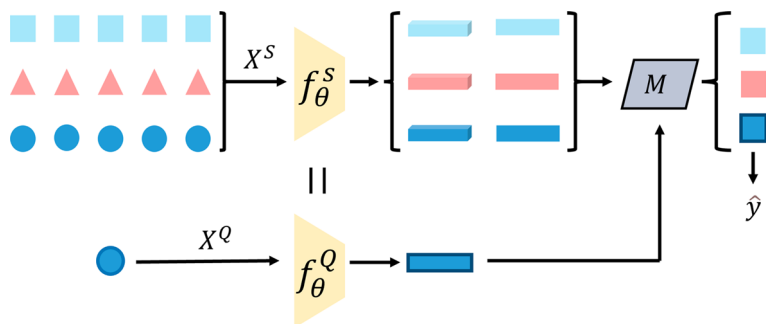
Fig. 9 Overview of the prototypical network (a 3-way 5-shot task for example)
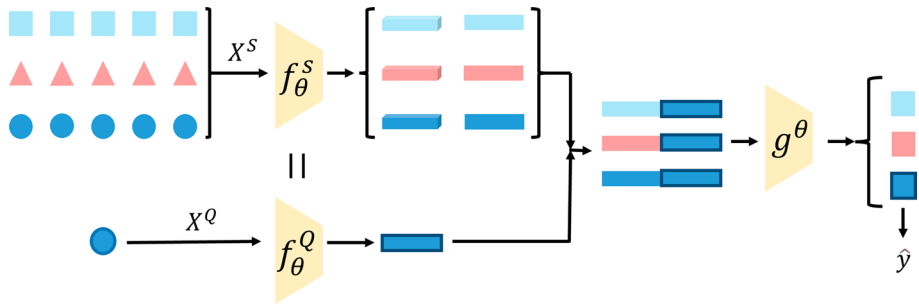
Fig. 10 Overview of the relation network (a 3-way 5-shot task for example)

For learning metrics, relation networks(Sung et al. 2018) are considered a classical approach. DFGR(Ma et al. 2020) first introduced relation networks into CSI human sensing, while Zhang et al. (2022c); Chen and Chang (2022)introduced graph convolution to further measure the relationships between different activity categories. Unlike matching networks and prototype networks, relation networks do not rely on distance functions to define metric values. Instead, they leverage neural networks to learn how to measure different features for recognition (as shown in Fig. 7.c). This aids in discovering relationships between features and improving the model's generalization ability. The structure of relation networks is illustrated in Fig. 10. The embedding module generates embedded features for query and support samples, and then the parameter metric module is used to determine whether they belong to the same category.

The main advantages of metrics-based few-shot learning methods are simplicity and strong generalization ability. Specifically, a metric that helps Wi-Fi sensing systems more accurately compare and distinguish between different signal patterns can be applied directly to a variety of new learning tasks simultaneously without fine-tuning. However, the assumption that the new learning task is similar to the task distribution in the previous training phase should be satisfied, otherwise the recognition rate will decrease. For example, in OneFi(Xiao et al. 2021), the difference between test and training positions is increased, and the recognition rate is 10% lower.

Fig. 11 The methods based on parameter optimization
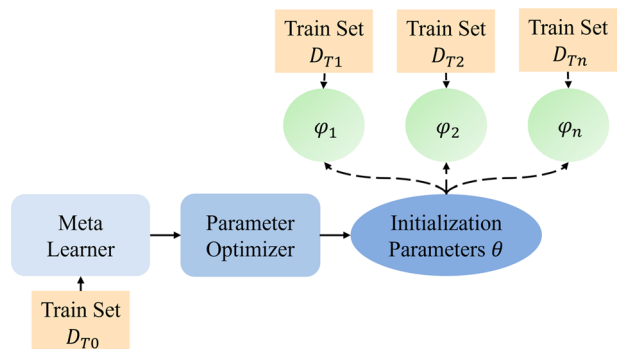
3.2.3 Meta-learning

The method based on meta-learning is generally understood as learning-to-learn, which refers to improving learning algorithms across multiple learning episodes. There are two main approaches based on meta-learning: meta-initialization and meta-optimizer.

Huang et al. (2022); Wang et al. (2021b); Gu et al. (2021); Zhang et al. (2022e); Gao et al. (2023); Owfi et al. (2023); Wei et al. (2023) aim to utilize meta-optimizer methods to achieve rapid parameter optimization in the case of small datasets. Existing work is based on meta-optimization and utilizes the model-agnostic meta-learning (MAML) (Finn et al. 2017) algorithm to learn the initial parameters of a model. The workflow of MAML is illustrated in Fig. 11. The model is expected to be trained using known initial parameters, and can rapidly converge to new tasks using only a small portion of the training data and a fixed number of iterations. After each iteration, better initial parameters can be obtained, enabling the base network to achieve high accuracy on new tasks with fewer updates. However, the MAML algorithm is sensitive to learning rates and requires extensive hyperparameter tuning. Additionally, optimizing initial parameters involves second-order derivatives, and computing second-order gradients is computationally expensive.

In addition to the aforementioned meta-learning approaches, there is another method of training a meta-optimizer, allowing the optimizer's parameters to be learned automatically. Ravi and Larochelle (2017) analyzed the drawbacks of traditional gradient update mechanisms in the few-shot scenario. They proposed the Meta-LSTM network for few-shot image classification and argued that conventional optimization algorithms based on gradient descent are not feasible for few-shot learning. In the Meta-LSTM architecture, LSTM serves as the meta-learner, while a deep convolutional neural network(CNN) functions as the primary learner. Through this approach, the meta-learner's acquired optimizer can rapidly converge the base model for each task.

Meta-learning-based methods can train meta-models suitable for multiple tasks, aiming to enable Wi-Fi sensing systems to quickly adapt to new tasks or environments. Meta-learning can also be integrated into a variety of classification, regression, and reinforcement learning models. However, meta-learning-based approaches may require longer training times to learn how to learn than the first two approaches. However, if there is mislabeled data, the accuracy of the method based on meta-learning implementation improves relative to metric learning(Zhang et al. 2022e).

To give readers a better understanding of the above, we briefly summarize the networks introduced above in Table 4.

4 Research in applying few-shot learning to Wi-Fi sensing

According to different application experiment settings and typical applications, CSI human sensing can be roughly divided into gesture recognition, activity recognition, positioning, user authentication, and crowd counting. This section gives an overview and summary of these applications of few-shot learning.

4.1 Experimental datasets

This paper collects and summarizes existing published CSI human sensing datasets for few-shot learning.

Table 4 A summary of presented few-shot learning approaches

Type	Approach	Typical network	Brief description
Transfer learning	Model-based		Replace the trained model with the original classification method and fine-tune it according to the scene
	Specifically designed strategies		Design different strategies to narrow the gap between different domains
Metric learning	Learning feature embeddings	Matching network	Use different networks to encode support and query inputs
	Learning class representations	Prototypical network	Propose class prototypes, and calculate the distances between query samples and class prototypes
	Learning metrics	Relation network	Employ a neural network to estimate a similarity score of feature embeddings
Meta-learning	Meta-optimizers	MAML	Enabling the model parameters to quickly adapt to new unknown tasks through gradient-based learning rules
	Meta-initialization	Meat-LSTM	Learn both the exact task-specific optimization of a classifier and good initialization values for the parameters of task-specific learner

(1) Widar3.0 (Zhang et al. 2022b): Widar3.0 aims to collect CSI signals from different domains, considering their impact on factors such as direction, location, environment, and person. This collection will facilitate the creation of two datasets: one for human-computer interaction gestures and another for digits 0 to 9. These datasets will be utilized in experiments to further investigate the influence of various factors on activity recognition. It was collected in a classroom, hall, and office with sixteen volunteers.

(2) SignFi (Ma et al. 2018): Before the SignFi work, most studies focused only on classifying simple gestures. However, this work has achieved the classification of nearly 300 commonly used gestures in daily life. The data were collected in the laboratory and classroom with five volunteer parameters for 10 or 20 repetitions of each action.

(3) ARIL (Wang et al. 2019): ARIL focuses on utilizing constructed neural networks to identify shared features of the same action in different positions, thereby having a small number of users in the dataset. The data were collected in a laboratory setting with a volunteer participating in 15 repetitions of each action.

(4) WIAR (Guo et al. 2019): The inconsistency of datasets has hindered comparison across related works in WIAR. Thus, WIAR has collected public activity datasets for both Wi-Fi-based and video-based human activity recognition, aiming to reduce labor and time costs while promoting the development of wireless sensing. The data were collected in three indoor environments with ten volunteers' participation; each action was repeated 30 times.

For more information on these datasets, refer to the corresponding references in Table 5. At present, most open-source datasets for CSI applications based on few-shot learning focus on human activity recognition. While there are open datasets available for localization and user authentication (Pan et al. 2023; Meneghello et al. 2023; Meng et al. 2023; Gassner et al. 2021), existing studies have not utilized them extensively, with many opting to create their own datasets instead.

In the following section, each work will be summarized briefly in a table, including the methods used, datasets employed, performance evaluation, as well as a concise overview of the dataset information such as the number of participants, categories of activities, environmental settings, number of locations, and data collection devices used.

4.2 Performance evaluation indicators

In this section, we will describe performance metrics. For classification tasks such as activity recognition, gesture recognition, and authentication, a common metric is

Table 5 Overview of public datasets for CSI applications based on few-shot learning

Dataset	Information					
	Application	Device	People	Classes	Environment	Locations
Widar3.0 (Niu et al. 2022)	Gesture	Intel 5300	16	16	3	5
SignFi (Ma et al. 2018)	Gesture	Intel 5300	5	276	2	1
ARIL (Wang et al. 2019)	Gesture	USRP	1	6	1	16
WIAR (Guo et al. 2019)	Activity	Intel 5300	10	16	6	4

Table 6 Common evaluation metrics

Evaluation metric	Formula
Classification accuracy (Accuracy)	$\text{Accuracy} = \frac{\text{Number of correctly classified samples}}{\text{Number of total samples}} \times 100\%$
Mean squared error (MSE)	$MSE = \frac{1}{n} \sum_{i=1}^{n} (fact_i - pred_i)^2$
Root mean square error (RMSE)	$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (fact_i - pred_i)^2}$
Cumulative distribution function (CDF)	$CDF(x) = \int_{-\infty}^{x} f(x)\, dx$

classification accuracy. From the point of view of positioning, it can be divided into location classification and location prediction. The former shares evaluation metrics with the classification task, while the latter primarily uses root mean square error (RMSE), mean square error (MSE), and cumulative distribution function (CDF) for evaluation. For cross-domain or new activity recognition, the dataset is usually divided according to the scenario, and the data in the test set never appears in the training set, such as training with data collected in the home environment and evaluating with data in the office environment. Different from traditional deep learning evaluation methods, this paper focuses on comparing the performance of 5-way 1-shot and 5-way 5-shot scenarios. Table 6 summarizes several evaluation metrics.

Because the CSI human sensing work based on few-shot learning is different from traditional image classification and other works, there is no unified data set for evaluation, and different working methods have different data sets, so it is impossible to make the same comparison under the same conditions. However, we summarize the performance of each work.

4.3 Application

4.3.1 Gesture recognition

Gesture recognition has become a hot research area in recent years. Gesture recognition technology can be widely used in virtual games, autonomous driving assistance systems, sign language recognition, and intelligent robot control. According to the commonness and characteristics of related work, the application of few-shot learning in gesture recognition is introduced from several aspects such as network transformation and acceleration calculation. In Table 7, we summarize the relevant work of CSI gesture recognition based on few-shot learning.

To improve the accuracy of the prototypical network, WiGr(Zhang et al. 2022f) introduces a new path to enhance the features. And orthogonal regularisation is introduced to increase the gap between different categories in the embedding space. DFGR (Ma et al. 2020) presents the relation network into CSI gesture recognition and uses the transferrable similarity evaluation ability to learn from the training set. Unlike the former, which uses cosine similarity to determine whether it is a class gesture, DFGR performs gesture recognition by training the classification network.

To speed up the CSI gesture recognition calculation, WiGR(Hu et al. 2021) introduced deep separable convolution and linear inverted residual structure to replace the original convolution block to reduce the calculation parameters and resource consumption. Compared with the traditional relationship network, the accuracy is 10% higher, and

Table 7 Application of few-shot learning in gesture recognition

Work	Method	Input type	Pre-process	Datasets	Accuracy
Zhang et al. (2022f)	Metric learning (Prototypical Net)	Amplitude Phase	Noise reduction	1.Widar3.0 2.ARIL 3.CSIDA	Different environment: Source: classroom; Target: office 1-shot Widar3.0: 60% CSIDA: 85% Different user: 1-shot Widar3.0: 78% CSIDA: 91% Different location: 1-shot Widar3.0: 59% ARIL: 56% Different orientation: 1-shot Widar3.0: 75%
Ma et al. (2020)	Metric learning (Relation Net)	Amplitude	Noise reduction	1. SignFi 2. Self-made (26 Classes, 5 Participants, 2 Environment, Intel 5300)	New classes: SignFi: Source: 100; classes Target: new 10 classes:97.6%, new 50 classes:89.5% Self: Source: 20 classes; Target: 6 new classes 1-shot 96.2% 5-shot 99.3%Different Environment: Source: lab; Target: home 5-shot: 99% Different User: 5-shot: 98%
Hu et al. (2021)	Metric learning (Relation Net)	Phase	Noise reduction	1. SignFi 2.CSIDA	New classes: SignFi: Source: 100 classes; Target: 10 new classes: 98.6% 20 new classes: 97.2% 30 new classes: 95.5% CSIDA: Source: 3 classes; Target: 2 new classes: 91.4% 3 new classes: 58.9% Different environment: SignFi: Source: Lab; Target: Home: 89% CSIDA: Source: Room1; Target: Room2: 89% Different user: SignFi: 92% CSIDA: 91% Different location: CSIDA: 90.5%

Table 7 (continued)

Work	Method	Input type	Pre-process	Datasets	Accuracy
Wang et al. (2022a)	Transfer learning			Self-made (40 Classes, 8 Participants, 4 Environment, Atheros)	New environment: Source: environment ABC; Target: environment D 94.14% Source: environment ABD; Target: environment C 93.89% Source: environment ACD; Target: environment B 93.99% Source: environment BCD; Target: environment A 94.40%
Yang et al. (2019)	Metric learning (Siamese Net)	Phase		Self-made (6 Classes, 10 Participants, 2 Environment, Atheros)	Different environment: Source: environment 1; Target: environment 2 65.3% Traditional deep learning 46.3% Source: environment 2; Target: environment 1 75.3% Traditional deep learning 48.7% Different user: Source: 5 users; Target: new 5 users 85.7% Traditional deep learning 53.1%
Xiao et al. (2021)	Transfer learning	Spectrogram	Noise reduction Data adaption Signal transform	Self-made (40 Classes, 10 Participants, 4 Environment, Intel 5300)	New classes: Different user: Source: participant 1; Target: other people 1-shot: 73.3% 5-shot: 94% Different environment: Source: environment 1; Target: other environment 1-shot: 80% 5-shot: 89.8% Different location Source: location 1; Target: other location 1-shot 70% 5-shot: 90% Different orientation: Source: orientation 1; Target: other orientation 1-shot: 78% 5-shot: 85%

the computational complexity is one-tenth of the original. Different from WiGR(Hu et al. 2021), which introduces lightweight convolution work, OneFi(Xiao et al. 2021) introduces a vision transformer(ViT) for feature extractors to realize parallel computing and reduce computing time. At the same time, researchers were inspired by the advancements in data enhancement techniques used in computer vision. Nonlinear optimization is applied to extract the body motion velocity information from multiple Doppler spectrograms of a specific pose. Each velocity component is then associated with its corresponding Doppler frequency component. The Doppler spectrogram of the transformed gesture can be generated by mapping the velocity components to the Doppler frequency components. This approach enriches the dataset with additional information for further analysis and research. After training, cosine similarity replaces the classification layer to fine-tune the feature extraction layer to recognize new gestures.

To reduce the difference between CSI gesture recognition domains, AirFi(Wang et al. 2022a) and Yang et al. (2019) introduce maximum mean discrepancy to confuse the gap between domains through domain alignment so that the difference between the embedding dimensions of the same action is slight. AirFi(Wang et al. 2022a) mainly addresses cross-environment problems. It adds Gaussian noise to increase the collected CSI samples and uses Laplace distribution and discriminator to reduce the dependence of the model on the source environment CSI to enhance the features. Yang et al. (2019) introduce the Siamese network to realize one-shot learning. While considering CSI as time information, Bi-LSTM is added to the feature extraction network to obtain time dimension features.

At present, there are some weaknesses in WiGR(Hu et al. 2021) and OneFi(Xiao et al. 2021). Recognition performance will be reduced when human activity is closer or further away from the receiving device. To achieve high accuracy in remote scenarios, we have to make additional efforts, such as improving the signal strength or the sensitivity of the receiver.

4.3.2 Activity recognition

In recent years, human activity recognition has received significant attention due to many potential applications that monitor human movement and behavior in indoor areas. Applications include health monitoring and fall detection for older people, context awareness, intelligent homes, and other IoT-based applications. We summarize here the activity recognition applications using few-shot learning.

FewSense(Yin et al. 2022) utilizes AlexNet as the feature extraction network. It incorporates an L2 normalization layer before the classification layer for feature normalization and aggregates embedded features from the same class while enlarging inter-class differences in features. The feature extraction network is then fine-tuned, and activity classification is achieved by using cosine similarity instead of the classification layer.

To accomplish the task of human activity recognition, WiLISensing(Ding et al. 2021), LI-HAR(Ding et al. 2022), ReWiS(Bahadori et al. 2022) and AutoFi(Yang et al. 2023) introduce prototypical networks to implement recognition tasks. WiLISensing(Ding et al. 2021) and Ding et al. (2022) focus on the generalization ability at different locations. ReWiS(Bahadori et al. 2022) mainly focuses on the environmental generalization ability and discusses the subsequent identification influence caused by the number of equipment antennas and transmitting and receiving frequencies. However, it uses only four activity categories. LI-HAR(Ding et al. 2022) based on WiLISensing(Ding et al. 2021), the CTS-AM

attention mechanism is introduced to improve feature extraction. ReWiS(Bahadori et al. 2022) extract the linear correlation between subcarriers by calculating the Pearson correlation coefficient. The influence of single and multiple receivers and sub-carrier resolution on activity recognition is also discussed.

Different from the previous classification tasks after signal preprocessing, Huang et al. (2022) and CSI-GDAM (Zhang et al. 2022c) hope to use a convolutional block attention module(CBAM) attention mechanism to remove the original amplitude signal noise. The former uses ResNet-9 as the backbone network to extract features and uses the meta-learning method for subsequent recognition. CSI-GDAM (Zhang et al. 2022c) is a method that aims to remove the difference between CSI activity sample feature vectors and the inner product. Based on this, it constructs node features and an adjacency matrix for the entire link graph. The components are updated using the difference between the inner product and the node feature vectors. Finally, the activity type is measured using graph convolution. In (Huang et al. 2022), meta-learning is utilized to tackle new activity and environmental problems. Additionally, to address the loss of time information during feature extraction by CNN, some studies introduce time coding to identify model parameters that are sensitive to task changes. Furthermore, an enhancement to the cross-entropy loss function is introduced to decrease the adverse effects of mislabeled data.

AFEE-MatNet (Shi et al. 2022) and Ding et al. (2021) introduced matching networks to complete activity recognition tasks. In contrast, Ding et al. (2021) found that the previous experiments ignored the impact of the initial state on the recognition results and used the matching network to overcome the effects of different initial conditions on CSI transmission. At the same time, Shi et al. (2022) combined activity-related feature extraction and enhancement methods with matching networks. The environmental noise irrelevant to the activity is filtered out, and the information related to the action is compressed and saved. Since human activities are not independent, a predictive detection and correction scheme is introduced to correct some classification errors that do not match the state transition of human behavior.

Like AirFi (Wang et al. 2022a), Wang et al. (2021b) introduce Wasserstein distance to accelerate convergence and improve the loss function to alleviate mode collapse. The virtual samples generated by FWGAN are trained by the method based on optimization (Finn et al. 2017). Unlike the former, LT-WIOB(Zhou et al. 2022) constructs triples to alleviate the urgent need for massive training data. The triple input is built from a small number of samples, and the intra-group dependence of the three inputs is measured. Then, a lightweight convolution block is introduced to reduce the amount of computation, and the loss function is optimized to improve the accuracy. AutoFi(Ma et al. 2018) introduces self-supervised learning and uses contrastive knowledge, mutual information, and geometric structure loss to keep the geometric structure of the two batch views consistent. After presenting the geometric self-supervised module, the average recognition rate is improved by 5%.

Table 8 summarises the related work on CSI activity recognition based on few-shot learning. CSI-GDAM (Zhang et al. 2022c) and Guo et al. (2019) hope to use the attention mechanism to replace the previous signal preprocessing. However, for few-shot learning, the signal fluctuations caused by new types of activities are different. If the attention mechanism parameters focus on the previous activity positions, the vital information about recent activities will be ignored. In addition to (Huang et al. 2022), there is no work considering the impact of wrong labels, which has a higher value for future practical applications.

Table 8 Application of few-shot learning in activity recognition

Work	Method	Input type	Pre-process	Datasets	Accuracy
Yin et al. (2022)	Transfer learning	Amplitude phase	Noise reduction	1.SignFi 2.Widar3.0 3.WiAR	New classes: 1-shot SignFi: Source: 200 classes; Target:76 new classes 57.9% (user s5) SignFi: Source: 150 classes; Target:25 new classes 71% (all users) Different Dataset(6-way 1-shot): Source: SignFi; Target: Widar 6 receiver: 75.5%
Bahadori et al. (2022)	Metric learning (Prototypical Net)	Amplitude phase	Data Adaption	Self-made (4 Classes, 1 Participant, 3 Environment, Nexmon)	Different Environment (4-way 5-shot): Source: Area1; Target: Area2 99.25% Area3 98.85% Different Device (4-way 5-shot): Source: Area1 20Mhz 3receiver 4 antennas; Target: Area2 20Mhz 1 receiver 1 antennas 59.75% 4 antennas 74.83% 80Mhz 1 receiver 1 antennas 78.25% 4 antennas 90.42% 20Mhz 3 receiver 1 antennas 96.5% 4 antennas 84.25% 80Mhz 1 receiver 1 antennas 92.75% 4 antennas 98.58%
Ding et al. (2022)	Metric learning (Prototypical Net)	Amplitude	Noise reduction	Self-made (4 Classes, 24 Locations; Intel 5300)	Different locations: (1-shot) Source: 4/6/8/12/24 locations; Target: 24 locations respectively 91.11% 92.23% 94.98% 96.00% 98.36% Different Subcarriers: (1-shot) Source: 4 locations; Target: 24 locations 90 subcarriers 91.11% 60 subcarriers 89.94 30 subcarriers 86.31% Different TX-RX antenna pairs(1-shot) Source: 4 locations; Target: 24 locations 1TX-1RX 86.31% 2TX-2RX 87% 3TX-3RX 85.6%

Table 8 (continued)

Work	Method	Input type	Pre-process	Datasets	Accuracy
Zhang et al. (2022c)	Metric learning (Relation Net)	Amplitude		1. SignFi 2. Self-made (16 Classes, 4 Participants, 2 Environment, Intel 5300)	New classes: SignFi: Source: 50 classes; Target: 50 new classes 5-way 1-shot:99.75% Self: Source: 8 classes; Target: 8 new classes 5-way 1-shot:93.12% Different Environment: SignFi: Source: home; Target: lab 5-way 1-shot:98.61% Self: Source: lab, Target: 8 conference 5-way 1-shot:92.71% New Classes and Different Environments: SignFi: Source: 50 classes lab; Target: 50 new classes home 5-way 1-shot:98.42% Self: Source: 8 classes lab; Target: 8 new classes conference 5-way 1-shot:91.32% Different Dataset: Source: SignFi; Target: Self 5-way 1-shot:98.42% 5-way-5shot:99.74%
Ding et al. (2021)	Metric learning (Matching Net)	Amplitude	Noise reduction	Self-made (4 Classes, 2 Environment, Intel 5300)	Different initial states: 1-shot:92.5% 3-shot 96.5%
Zhou et al. (2022)	Metric learning (Siamese Net)	Spectrogram	Noise reduction Data adaption Signal transform	Self-made (8 Classes, 10 Participants, 2 Environment, Intel 5300)	Diverse Individuals 1-shot 87.3% 5-shot 91.3% Diverse environment 1-shot 86.6% 5-shot 90.4% Different input and device: 1 antenna: PAC 75% FFT 78% STFT 92% 2 antenna: PAC 79% FFT 82% STFT 93% 3 antenna: PAC 80.0% FFT 84% STFT 94%

Table 8 (continued)

Work	Method	Input type	Pre-process	Datasets	Accuracy
Huang et al. (2022)	Meta-learning	Amplitude		1. SignFi 2. Self-made (16 Classes, 4 Participants, 2 Environment, Intel 5300)	New classes: SignFi: Source 50 classes; Target: 50 new classes 1-shot: Area1: 96, Area2 95% Different environment: SignFi: Source: lab 50 classes; Target: home 50 classes 5-way 1-shot: 85.3% 5-way 5-shot: 87.4% Self: Source: lab 8 classes; Target: conference 8 classes 5-way 1-shot:76.5% 5-way 5-shot: 80.5%
Yang et al. (2023)	Metric learning (Prototypical Net)	Amplitude	Noise reduction	Self-made (40 Classes, 8 Participants, 4 Environment, Atheros)	New Classes: Widar: 10-shots: 55.60% 20-shots: 63.8% UTHAR: 10-shots: 66.8% 20-shots 77.8% Self: avg 86.83%
Zhang et al. (2022e)	Meta-learning	Amplitude		1.SignFi 2.Self-made (16 classes, 4 participants, 2Environment, Intel 5300)	New classes: SignFi: 5-way 5-shot:84.4% 5-way 1-shot:65.7% Different Environment Self: 5-shot: 93.5% 1-shot: 68.2%
Shi et al. (2022)	Metric learning (Matching Net)	Spectrogram	Noise reduction Data adaption Signal transform	Self-made (6 Classes, 5 Participants, 6 Environment, Intel 5300)	Different Environment: 6 way 81.6% Different User: 6 way 92%
Ding et al. (2021)	Metric learning (Prototypical Net)	Amplitude	Noise reduction	1.UTHAR 2.Widar3.0 3.Self-made (8 Classes, 14 Participants, 2 Environment, Atheros)	Different locations: Source: 4/6/8/12/24 locations; Target: 24 locations respectively 90.92% 92.75% 93.88% 95.92% 98.48%

Table 8 (continued)

Work	Method	Input type	Pre-process	Datasets	Accuracy
Wang et al. (2021b)	Meta-learning	Amplitude	Noise reduction	1.SignFi 2. Self-made (16 Classes, 4 Participants, 4 Environment, Intel 5300)	New Classes: SignFi: Source: 100 classes; Target: 100 new classes 1-shot: 90% 5-shot: 95% Self: Source: 8 classes; Target: 8 new classes 1-shot: 81.9% 5-shot: 91.5% Different User: SignFi: Source: user 1; Target: user 2/3/4 1-shot: 83.6% 5-shot: 93.5% Self: Source: user 1; Target: user 2/2/4/5 1-shot: 92.5% 5-shot: 98.1% Different Environment: SignFi: Source: lab; Target: home (20 classes) 1-shot:98.7% 5-shot: 99% SignFi: Source: home; Target: lab (20 classes) 1-shot: 94% 5-shot: 97.9% Self: 1-shot:83.3% 5-shot: 92.6% Different Environment and new Classes: SignFi: Source: lab, 100 classes Target: home 100 new classes 1-shot: 91% 5-shot: 97.5% Self: Source: lab 8 classes; Target: Conference 8 new classes 1-shot: 80.1% 5-shot: 90.5%

4.3.3 Location

Location-based services are ubiquitous and indispensable in our daily lives. In outdoor positioning, GPS is a very effective positioning method. However, in indoor environments, due to the influence of building occlusion, GPS signals will be disturbed and cannot provide accurate positioning. Wi-Fi positioning technology has become an important research direction in indoor positioning because of its simple equipment, high communication efficiency, and comprehensive coverage.

LESS (Zhang et al. 2023a) introduces the Wasserstein generative adversarial neural network (GAN) to extend the sparsely collected fingerprints and constructs a relationship network to calculate the location information of local proximity in the low-dimensional manifold space. CSI-MML (Wang et al. 2024) uses a prototypical network, introduces a CBAM attention mechanism to extract features, and applies multi-scale metric learning to measure the consistency of data distribution and the difference in local feature similarity between samples. The similarity between samples is effectively measured by measuring the global and local similarities between samples. Chen and Chang (2022) introduce graph networks as in (Zhang et al. 2022c). Firstly, the CNN is used for feature extraction to construct the input of a graph network. Then, inter-class samples are implicitly constructed, and graph convolution is used to update the relationship between intra-class samples.

MAML (Owfi et al. 2023) is used to complete few-shot learning, and MMD is introduced to reduce the weight of a given training task depending on the difference between the source task and the target task. Different from the former work, for the gap between environment and task, MetaLoc (Gao et al. 2023) proposed MAML-TS and MAML-DG based on MAML to complete the localization in the new environment. MAML-TS uses MMD to discover the best environment-specific parameters regarding task similarity. In the MAML-DG paradigm, modifying the loss function forces the loss in different training environments to decrease in similar directions, enabling faster convergence and better adaptation of the learned meta-parameters. Owfi et al. (2023) proposed TB-MAML to solve the persistent problem of vague generalization in traditional trained DL-based localization models and improved the generalization in the case of limited data sets.

Table 9 Application of few-shot learning in localization

Work	Method	DataSet	Accuracy
Wei et al. (2023)	Meta-learning	4 Environment, 20 Location; Intel 5300	RMSE: A: 1.09m B: 1.13m C: 1.12m D: 1.97m
Gao et al. (2023)	Meta-learning	2 Environment, 90 Location; Nexmon	Mean Error: A: 2.10m B: 3.10m
Wang et al. (2024)	Transfer learning	1Environment, 20 Location; 2 path Intel 5300	Mean Error: Path1 1.47m Path2 1.43m
Zhang et al. (2023a)	Metric learning	2 Environment, 460 Location; Intel 5300	Mean Error: 1.52m
Chen and Chang (2022)	Metric learning (Relation Net)	1 Environment, 16 Location; 1 Environment, 18 Location	A to B avg 73% B to A avg 86%
Owfi et al. (2023)	Meta-learning	33 scenarios 12 Location; Intel 5300	Mean Error: 50 cm

Table 9 summarises the related work on CSI human location based on few-shot learning. The existing few-shot learning works on CSI positioning all solve the cross-domain problem of fingerprint-based positioning. Due to the diverse variations in CSI signals across different environments, traditional deep learning methods face challenges in training models in the target domain, especially when there are insufficient samples available. In this case, the trained model cannot maintain recognition accuracy in the new environment.

4.3.4 Other

In addition to activity recognition and location, some studies have applied few-shot learning to crowd counting and user authentication. Office environments such as air conditioning can be controlled automatically by obtaining accurate numbers of people. However, the accuracy rate of the trained model decreases seriously after switching the environment. For example, it is pointed out in (Hou et al. 2023) that after applying the deep learning model trained in the office environment to a more spacious conference room, the accuracy rate drops from 99% to 12%. It seriously hinders large-scale deployment and is a suitable method for few-shot learning.

DASECount (Hou et al. 2023; Hou et al. 2022) adopt the same method to conduct classification model training in the source domain and introduce a logistic regression classifier to replace the training classifier to complete the classification. The difference is that DASE-Count (Hou et al. 2023) improved (Hou et al. 2022) and extracted the amplitude and phase features, respectively. Later, to improve the generalization ability of the feature extractor, knowledge distillation was introduced.

ResMon (Zheng et al. 2023b) introduces few-shot learning into the respiratory detection system. In contrast to traditional respiratory rate detection, it focuses more on the detection of respiratory states such as stable breathing and coughing. Unlike the introduction of lightweight convolution or attention mechanisms in the above-mentioned activity recognition, it incorporates Bayesian neural networks to address the overfitting issue in traditional CNN and introduces Kullback–Leibler (KL) divergence to approximate the true posterior probability.

Traditional user authentication relies on human biometric features such as iris scans, fingerprints, etc. However, in wireless sensing, the crucial technique is capturing and analyzing each individual's unique biological motion characteristics, including factors like gait and hand gestures. The variations in signals caused by different user movements can be leveraged for authentication purposes.

WiONE (Gu et al. 2021) implements user authentication by relying on a handwritten password. Unlike traditional filtering to extract meaningful information, a behavior enhancement model based on Rician fading is designed to improve the quantitative model of human behavior response sensitivity. The prototypical network is used to realize one-shot user authentication in the same environment, which is more concerned with using fewer data to achieve a higher recognition rate.

Different from WiONE (Gu et al. 2021), CAUTION (Wang et al. 2022b), and Meta-GanFi (Zhang et al. 2022d) utilize gait features to complete user identification. CAUTION (Wang et al. 2022b) uses prototypical networks for gait recognition and intrusion detection. The distance between the new quality and the two nearest points is calculated, and the threshold is constantly optimized to realize intrusion detection. MetaGanFi (Zhang et al. 2022d) proposes a conditional cycle-consistent gait GAN model to learn the mapping

Table 10 Application of few-shot learning in other applications

Work	Method	Application	Input Type	Datasets	Accuracy
Hou et al. (2022)	Transfer learning	Crowd counting	Amplitude	2 Environment, 4 Scenarios, 9 Participants, Atheros	Different Scenario: Source: Scenario A; Target: Scenario B 1-shot: 71.33% 5-shot 86.73% Scenario C 1-shot: 78.85% 5-shot 94.18% Scenario D 1-shot: 79.24% 5-shot 92.13%
Hou et al. (2023)	Transfer learning	Crowd counting	Amplitude phase	2 Environment, 4 Scenarios, 9 Participants, Atheros	Different Scenarios: Source: Scenario A; Target: Scenario B 1-shot:80.65% 5-shot:96.85% Scenario C 1-shot:77.17% 5-shot:94.61% Scenario D 1-shot:76.26% 5-shot:92.68% Different Input: Source: Scenario A; Target: Scenario C Amplitude:1-shot:67.56% 5-shot:88.20% phase features:1-shot:59.04% 5-shot:84.91% all:1-shot:76.26% 5-shot:92.68%
Gu et al. (2021)	Metric learning (Prototypical Net)	User authentication	Amplitude	2 Environment, 20 Participants, (16 legal, 3 ilegal) Intel 5300	Different Environment: Source: Area1; Target: Area1: ≈94% Area2: ≈88% Area3: ≈ 87% Area1: ≈91% Area2: ≈96% Area3: ≈95% Area1: ≈91% Area2: ≈94% Area3: ≈95% Different Password: Simple password: 1-shot: ≈94% 5-shot: ≈96% Complex password: 1-shot: ≈91% 5-shot: ≈93% Different Input: Energy input signals: 1-shot: ≈<90% 5-shot: ≈91% Spectrogram: 1-shot: ≈92% 5-shot: ≈93% All: 1-shot: ≈94% 5-shot: ≈96%
Wang et al. (2022b)	Metric learning (Prototypical Net)	User authentication Intruder Detection	Amplitude	2 Environment, 3 Scenarios, 20 Participants (15 legal 5 illegal), Atheros	Different surrounding disturbance: Source: Scenario A; Target: Scenario B User Identification: Area1: 85.98% Area2: 85.06% Intruder Detection: Area1: 92% Area2: 90% Different Users' dressing: Source: Scenario A; Target: Scenario C Area1: 86.29% Area2: 87.36%

Table 10 (continued)

Work	Method	Application	Input Type	Datasets	Accuracy
Zhang et al. (2022d)	Meta-learning	User authentication	Amplitude phase	2 Environment, 11 Participants 4 Trajectories, (1 transmitter 6 receivers)	7 participants for training, 4 participants for test Different trajectory: Source: 1/3 trajectory; Target: 2/4 trajectory 87.25% 90.50% 93.50% 93.75% for 1, 3, 5, 7-shot Different Receiver: Source: 1/3/5 receiver; Target: 2/4/6 receiver 86.84% 90.16% 91.25% 93.42% for 1, 3, 5, 7-shot
Zheng et al. (2023b)	Meta-learning	Respiration state	Amplitude phase	2 Environment, 6 Participants Atheros	Different User(4-way 5-shot): Source: Area1; Target: Area2 86.5% Area3 83.8% Different Environment(4-way 1-shot): Source: Area1; Target: Area2 89.8% Area3 96.11% Different Location(4-way 5-shot): Source: Area1; Target: Area1 0.5m: 96.22% 1.5m: 85.12% 2.5m: 83.11% (Traditional deep learning 82.22% 72.50% 56.67%) Different Device(4-way 5-shot): Source: Area1(2.4GHz, 20MHz); Target: Area2 (5GHz, 40MHz) 95% (Traditional deep learning 63.38%)

between multiple domains, which then acts as a domain filter that converts the multi-domain CSI into the single-domain CSI.

Table 10 summarises the related work on CSI crowd counting and user authentication based on few-shot learning. Compared to activity recognition and localization tasks, there is relatively less work done on user authentication and crowd counting in the context of few-shot learning. Moreover, there is a scarcity of open-source datasets available for these tasks compared to others.

4.3.5 Discussion

In the above section, an introduction to various CSI-based human sensing applications using few-shot learning was presented, and the related work was summarized in tables.

Six works (Hou et al. 2022; Huang et al. 2022; Zhang et al. 2022e; Wang et al. 2022a; Yang et al. 2019) and (Zhang et al. 2022c) omitted the signal processing module, they used raw amplitude and phase as inputs. CSI-GDAM (Zhang et al. 2022c), CSI-MML (Wang et al. 2024), LI-HAR (Ding et al. 2022), and (Huang et al. 2022) aimed to enhance accuracy by leveraging attention mechanisms for denoising. However, in the visual domain, some related studies (Hou et al. 2019) have raised concerns about the inability of traditional attention mechanisms to adapt to new categories.

From the tables mentioned above, most works utilize amplitude as the network input due to the stability of amplitude compared to phase. WiGr (Zhang et al. 2022f) discussed the influence of amplitude and phase on recognition accuracy, but due to the small magnitude of gesture activities, phase changes are more suitable compared to amplitude. Consequently, phase has a higher recognition accuracy than amplitude, but the improvement in accuracy varies across different datasets. In contrast to the above studies, DASECount(Hou et al. 2023) combines layer normalization of amplitude and phase difference as input. Compared to some applications using amplitude and phase separately as inputs, it achieved higher accuracy.

Although signal processing and feature engineering introduce overhead, in certain cases, the processed features can enhance sensing accuracy. ResMon (Zheng et al. 2023b) reduces the impact of frequency offset by utilizing the CSI ratio, and subsequently alters the relative proportions of the original real and imaginary parts of the CSI (Zeng et al. 2021) as inputs. A discussion comparing data with and without filtering is conducted, revealing that the original CSI signals may not meet the practical application requirements. WiONE (Gu et al. 2021) utilizes Rician fading to enhance the variation of CSI. Additionally, experimental results confirm the impact of energy images and spectrogram on recognition rate, with the spectrogram recognition rate surpassing that of energy images. Combining these two methods results in higher recognition rates than using either input alone. Compared to directly converting CSI into spectrograms, AFEE-MatNet (Shi et al. 2022) adopts the AFEE mechanism, which reduces the size of the input signal's CSI matrix, thereby reducing training time.

Usually, we can observe that for coarse-grained human actions, such as walking, the amplitude is suitable as input, while for fine-grained actions, such as gestures, the phase is more appropriate. Combining both amplitude and phase as inputs can further improve the recognition rate. Zhou et al. (2022) and Xiao et al. (2021) convert CSI into a spectrogram, which varies significantly in amplitude and phase compared to the traditional time domain,

but both of these operations perform more preprocessing on the signal. Currently, concatenation is often used for input fusion, but more advanced fusion methods at a deeper level are not considered. For example, UniFi (Liu et al. 2024) utilizes a self-attention mechanism to achieve the fusion of multiple devices and multiple types of inputs. In addition, performing preprocessing on input signals can also enhance the recognition rate and accelerate training speed.

Currently, there are three main methods in few-shot learning: transfer learning, metric learning, and meta-learning approaches. Compared to the former two, meta-learning has lower efficiency as it requires retraining and parameter tuning on the support set, yet it offers greater flexibility. Transfer learning methods can directly use model parameters or adapt to new tasks with limited data through fine-tuning. The former tends to be more efficient, while the latter is less so, exemplified by FewSense (Yin et al. 2022), which sees an enhanced recognition rate post-fine-tuning. Metric learning focuses on learning the similarity measure between samples and can directly apply trained model parameters to a new domain, with efficiency largely depending on subsequent metric strategies.

Regarding scalability and complexity, transfer learning methods typically offer simplicity and directness by leveraging pre-trained models for fine-tuning novel tasks. Conversely, metric learning approaches may entail more intricate distance measurement and similarity calculations between samples, increasing methodological complexity, as seen in CSI-MML (Wang et al. 2024) where global and local similarities are employed. CSI-GDAM (Zhang et al. 2022c) introduces graph convolution for feature measurement. In contrast, meta-learning methods may demand greater computational resources and time for parameter retraining and adjustment on support sets, potentially limiting their efficacy with large-scale datasets or real-time applications. Additionally, meta-learning approaches often leverage techniques such as GANs to enhance classification accuracy. For instance, (Wang et al. 2021b) introduced in FWGAN has shown improvements in classification accuracy ranging up to 20%.

5 Issues and future challenges

The previous human sensing based on CSI relied on pre-set or trained models, and the training and performance largely depended on the availability of a sufficient amount of labeled data. It may not effectively identify activities in new domains, which means that maintaining recognition accuracy was difficult. Therefore, the few-shot learning is an important solution to this challenge. The problem of cross-domain Wi-Fi sensing has been identified as a key challenge in the field of Wi-Fi sensing (Wang et al. 2021a; He et al. 2020; Wang et al. 2018b). How to maintain model accuracy in the target domain dataset has been a crucial research topic in recent years, and few-shot learning has played a significant role in this regard.

Despite progress made in CSI human sensing through few-shot learning, obstacles such as limited access to CSI-related open-source datasets and the need for further exploration of multimodal fusion techniques have impeded further advancements. Overcoming these challenges is critical to promoting the development of new applications and advancing the field. Therefore, several key challenges must be addressed to promote the development of new applications and further advance the field.

5.1 Across multiple domains

Most of the current cross-domain problems of few-shot CSI human activity recognition only focus on simple single-domain (only cross-environment, only cross-user, only cross-location) problems and do not extend to multi-domain scenarios. At the same time, besides MetaLoc (Gao et al. 2023) and Hou et al. (2023), the current research on CSI human sensing based on few-shot learning only focuses on line-of-sight communication scenarios, and the cross-domain of non-line-of-sight communication scenarios is still not wholly concerned.

5.2 Multiple devices

As seen from the tables in section 4, most current related research is based on data collected from the same devices. FewSense (Yin et al. 2022) conducted experiments on different data sets of the same equipment. With more and more tools being released, such as Nexmon (Gringoli et al. 2019), PicoScences-Wi-Fi (Jiang et al. 2022), Esp32CSIToolkit (Hernandez and Bulut 2020), etc, we should consider more data from different devices. At the same time, as mentioned in (Cominelli et al. 2023), a relatively limited number of open-source datasets are available. Collecting large amounts of CSI data and preparing data sets that are easy to annotate is a tedious task that requires specific software tools and repeated activities multiple times. Compared with traditional computer vision work, most current work datasets are self-built and not open source, hindering the reproducibility of research results.

5.3 Multiple applications

Most of the current work on CSI human sensing based on few-shot learning focuses on activity recognition. There are still a few applications in other fields, such as respiration/heart rate estimation. Currently, more and more deep learning-based works focus on Wi-Fi imaging (Yu et al. 2022) and Wi-Fi-based human pose estimation (Zhou et al. 2022; Wang et al. 2022c; Yang et al. 2022a). At the same time, Yang et al. (2022b) use the few-shot learning method based on meta-learning to overcome the environmental impact of radio frequency identification (RFID) and complete human pose estimation in different scenes.

5.4 Robustness for roughly labeled samples

Considering the influence of human labeling error during data collection, there are usually some inaccurately labeled samples in practical scenarios. However, most existing few-shot learning techniques cannot deal with noisy labels, as they are based initially on being equipped with ideal data collection conditions and accurate labels. Due to the heavy dependence on accurate supervision information, few-shot learning algorithms are easily disturbed by irrelevant noise features, resulting in poor learning results. Therefore, improving the algorithm's robustness is necessary to learn from roughly labeled samples.

6 Conclusion

This paper conducts a comprehensive review of the application of few-shot learning in the context of CSI-based human sensing. Initially, it introduces the concept of CSI alongside traditional signal processing techniques, highlighting the necessity to address the challenges posed by cross-domain sensing. Subsequently, few-shot learning is explored and categorized based on different methods of implementation, with a discussion on the strengths and weaknesses associated with each approach. Furthermore, the paper compares the current applications in this field, identifying several areas that warrant further investigation, such as cross-modality and cross-device compatibility, in future studies. The primary objective of this review is to provide readers with a clear understanding of the prevailing research landscape concerning the use of few-shot learning for human sensing with CSI. Since the integration of few-shot learning into CSI-based human sensing is still in the early stages, requiring more advancements, this review serves as a valuable resource for researchers seeking a comprehensive overview of this emerging field.

Data availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare no Conflict of interest.

References

Abdelnasser H, Youssef M, Harras KA (2015) Wigest: a ubiquitous wifi-based gesture recognition system. In: 2015 IEEE Conference on Computer Communications (INFOCOM), pp. 1472–1480. https://doi.org/10.1109/INFOCOM.2015.7218525

Antonelli S, Avola D, Cinque L, Crisostomi D, Foresti GL, Galasso F, Marini MR, Mecca A, Pannone D (2022) Few-shot object detection: a survey. ACM Comput Surv (CSUR) 54(11s):1–37. https://doi.org/10.1145/3519022

Bahadori N, Ashdown J, Restuccia F (2022) Rewis: Reliable wi-fi sensing through few-shot multi-antenna multi-receiver csi learning. In: 2022 IEEE 23rd International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM), pp. 50–59. https://doi.org/10.1109/WoWMoM54355.2022.00027

Chen B-J, Chang RY (2022) Few-Shot Transfer Learning for Device-Free Fingerprinting Indoor Localization. Paper presented at the ICC 2022 - IEEE International Conference on Communications, pp. 4631–4636. https://doi.org/10.1109/ICC45855.2022.9839217

Chen C, Zhou G, Lin Y (2023) Cross-domain wifi sensing with channel state information: a survey. ACM Comput Surv 55(11):1–37. https://doi.org/10.1145/3570325

Cominelli M, Gringoli F, Restuccia F (2023) Exposing the csi: A systematic investigation of csi-based wi-fi sensing capabilities and limitations. In: 2023 IEEE International Conference on Pervasive Computing and Communications (PerCom), pp. 81–90. https://doi.org/10.1109/PERCOM56429.2023.10099368

Ding X, Jiang T, Zhong Y, Wu S, Yang J, Xue W (2021) Improving wifi-based human activity recognition with adaptive initial state via one-shot learning. In: 2021 IEEE Wireless Communications and Networking Conference (WCNC), pp. 1–6. https://doi.org/10.1109/WCNC49053.2021.9417590

Ding X, Jiang T, Zhong Y, Yang J, Huang Y, Li Z (2021) Device-free location-independent human activity recognition via few-shot learning. In: 2021 IEEE/CIC International Conference on Communications in China (ICCC Workshops), pp. 106–111. https://doi.org/10.1109/ICCCWorkshops52231.2021.9538898

Ding X, Jiang T, Zhong Y, Wu S, Yang J, Zeng J (2022) Wi-fi-based location-independent human activity recognition with attention mechanism enhanced method. Electronics. https://doi.org/10.3390/electronics11040642

Duan R, Li D, Tong Q, Yang T, Liu X, Liu X (2021) A survey of few-shot learning: an effective method for intrusion detection. Secur Commun Netw 2021:1–10. https://doi.org/10.1155/2021/4259629

Fe-Fei L, Fergus Perona (2003) A bayesian approach to unsupervised one-shot learning of object categories. In: Proceedings Ninth IEEE International Conference on Computer Vision, pp. 1134–11412. https://doi.org/10.1109/ICCV.2003.1238476

Finn C, Abbeel P, Levine S (2017) Model-agnostic meta-learning for fast adaptation of deep networks. In: Proceedings of the 34th International Conference on Machine Learning - Volume 70. ICML'17, pp. 1126–1135. JMLR.org, Sydney NSW Australia. https://doi.org/10.5555/3305381.3305498

Gao R, Zhang M, Zhang J, Li Y, Yi E, Wu D, Wang L, Zhang D (2021) Towards position-independent sensing for gesture recognition with Wi-Fi. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 5(2):1–28. https://doi.org/10.1145/3463504

Gao J, Wu D, Yin F, Kong Q, Xu L, Cui S (2023) Metaloc: learning to learn wireless localization. IEEE Journal on Selected Areas in Communications, 1–1 https://doi.org/10.1109/JSAC.2023.3322766

Gassner A, Musat C, Rusu A, Burg A (2021) OpenCSI: An open-source dataset for indoor localization using CSI-based fingerprinting. arXiv e-prints, 2104–07963. https://doi.org/10.48550/arXiv.2104.07963arXiv:2104.07963 [eess.SP]

Gringoli F, Schulz M, Link J, Hollick M (2019) Free your csi: A channel state information extraction platform for modern wi-fi chipsets. In: Proceedings of the 13th International Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization, pp. 21–28. https://doi.org/10.1145/3349623.3355477

Gringoli F, Cominelli M, Blanco A, Widmer J (2022) Ax-csi: Enabling csi extraction on commercial 802.11 ax wi-fi platforms. In: Proceedings of the 15th ACM Workshop on Wireless Network Testbeds, Experimental Evaluation & CHaracterization, pp. 46–53. https://doi.org/10.1145/3477086.3480833

Gu Y, Yan H, Dong M, Wang M, Zhang X, Liu Z, Ren F (2021) Wione: one-shot learning for environment-robust device-free user authentication via commodity wi-fi in man-machine system. IEEE Trans Comput Soc Syst 8(3):630–642. https://doi.org/10.1109/TCSS.2021.3056654

Gu Y, Zhang X, Wang Y, Wang M, Yan H, Ji Y, Liu Z, Li J, Dong M (2022) Wigrunt: wifi-enabled gesture recognition using dual-attention network. IEEE Trans Human-Mach Syst 52(4):736–746. https://doi.org/10.1109/THMS.2022.3163189

Gu H, Yang J, Gui G, Gacanin H (2023) Triplet matchnet based indoor position method using CSI fingerprint similarity comparison. IEEE Trans Veh Technol 72(12):16905–16910. https://doi.org/10.1109/TVT.2023.3289631

Guo L, Wang L, Lin C, Liu J, Lu B, Fang J, Liu Z, Shan Z, Yang J, Guo S (2019) Wiar: a public dataset for wifi-based activity recognition. IEEE Access 7:154935–154945. https://doi.org/10.1109/ACCESS.2019.2947024

Gupta N, Gupta SK, Pathak R (2022) Human activity recognition in artificial intelligence framework: a narrative review. Artif Intell Rev 55:4755–4808. https://doi.org/10.1007/s10462-021-10116-x

Halperin D, Hu W, Sheth A, Wetherall D (2011) Tool release: gathering 802.11 n traces with channel state information. ACM SIGCOMM Comput Commun Rev 41(1):53–53. https://doi.org/10.1145/1925861.1925870

He Y, Chen Y, Hu Y, Zeng B (2020) Wifi vision: sensing, recognition, and detection with commodity mimo-ofdm wifi. IEEE Internet Things J 7(9):8296–8317. https://doi.org/10.1109/JIOT.2020.2989426

Hernandez SM, Bulut E (2020) Lightweight and standalone iot based wifi sensing for active repositioning and mobility. In: 2020 IEEE 21st International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM), pp. 277–286. https://doi.org/10.1109/WoWMoM49955.2020.00056

Hou R, Chang H, Ma B, Shan S, Chen X (2019) Cross attention network for few-shot classification. In: Proceedings of the 33rd International Conference on Neural Information Processing Systems. Curran Associates Inc., Red Hook, NY, USA. https://doi.org/10.5555/3454287.3454647

Hou H, Bi S, Zheng L, Lin X, Quan Z (2022) Sample-efficient cross-domain wifi indoor crowd counting via few-shot learning. In: Proceedings of the 2022 31st Wireless and Optical Communications Conference (WOCC), pp. 132–137. https://doi.org/10.1109/WOCC55104.2022.9880570

Hou H, Bi S, Zheng L, Lin X, Wu Y, Quan Z (2023) Dasecount: domain-agnostic sample-efficient wireless indoor crowd counting via few-shot learning. IEEE Internet Things J 10(8):7038–7050. https://doi.org/10.1109/JIOT.2022.3228557

Hu P, Tang C, Yin K, Zhang X (2021) Wigr: a practical wi-fi-based gesture recognition system with a lightweight few-shot network. Appl Sci. https://doi.org/10.3390/app11083329

Huang S, Chen Y, Wu D, Yu G, Zhang Y (2022) Few-shot learning for human activity recognition based on CSI. https://doi.org/10.1109/CACML55074.2022.00074

Huang G, Laradji I, Vázquez D, Lacoste-Julien S, Rodríguez P (2023) A survey of self-supervised and few-shot object detection. IEEE Trans Pattern Anal Mach Intell 45(4):4071–4089. https://doi.org/10.1109/TPAMI.2022.3199617

Jiang Z, Luan TH, Ren X, Lv D, Hao H, Wang J, Zhao K, Xi W, Xu Y, Li R (2022) Eliminating the barriers: demystifying wi-fi baseband design and introducing the picoscenes wi-fi sensing platform. IEEE Internet Things J 9(6):4476–4496. https://doi.org/10.1109/JIOT.2021.3104666

Khamis A, Kusy B, Chou CT, Hu W (2020) Wirelax: towards real-time respiratory biofeedback during meditation using wifi. Ad Hoc Netw 107:102226. https://doi.org/10.1016/j.adhoc.2020.102226

Khan NS, Ghani MS (2021) A survey of deep learning based models for human activity recognition. Wireless Pers Commun 120(2):1593–1635. https://doi.org/10.1007/s11277-021-08525-w

Khan A, Sohail A, Zahoora U, Qureshi AS (2020) A survey of the recent architectures of deep convolutional neural networks. Artif Intell Rev 53:5455–5516. https://doi.org/10.1007/s10462-020-09825-6

Koch G, Zemel R, Salakhutdinov R et al (2015) Siamese neural networks for one-shot image recognition. In: ICML Deep Learning Workshop, vol. 2. Lille

Li H, Yang W, Wang J, Xu Y, Huang L (2016) Wifinger: talk to your smart devices with finger-grained gesture. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 250–261. https://doi.org/10.1145/2971648.2971738

Liu Y, Zhang H, Zhang W, Lu G, Tian Q, Ling N (2022) Few-shot image classification: current status and research trends. Electronics. https://doi.org/10.3390/electronics11111752

Liu Y, Yu A, Wang L, Guo B, Li Y, Yi E, Zhang D (2024) Unifi: a unified framework for generalizable gesture recognition with wi-fi signals using consistency-guided multi-view networks. Proc ACM Interact Mob Wearable Ubiquitous Technol. https://doi.org/10.1145/3631429

Lu J, Gong P, Ye J, Zhang J, Zhang C (2020) A survey on machine learning from few samples. Preprint at https://ui.adsabs.harvard.edu/abs/2020arXiv200902653L. https://doi.org/10.48550/arXiv.2009.02653

Ma Y, Zhou G, Wang S, Zhao H, Jung W (2018) Signfi: sign language recognition using wifi. Proc ACM Interact Mob Wearable Ubiquitous Technol. https://doi.org/10.1145/3191755

Ma Y, Zhou G, Wang S (2019) Wifi sensing with channel state information: a survey. ACM Comput Surv (CSUR) 52(3):1–36. https://doi.org/10.1145/3310194

Ma X, Zhao Y, Zhang L, Gao Q, Pan M, Wang J (2020) Practical device-free gesture recognition using wifi signals based on metalearning. IEEE Trans Industr Inf 16(1):228–237. https://doi.org/10.1109/TII.2019.2909877

Meneghello F, Fabbro ND, Garlisi D, Tinnirello I, Rossi M (2023) A CSI dataset for wireless human sensing on 80 MHz wi-fi channels. IEEE Commun Mag 61(9):146–152. https://doi.org/10.1109/MCOM.005.2200720

Meng X, Li W, Zhao Z, Liu Z, Wang H (2023) Hierarchical contrastive learning for csi-based fingerprint localization. In: International Conference on Artificial Neural Networks, pp. 306–318. https://doi.org/10.1007/978-3-031-44198-1_26 . Springer

Niu K, Zhang F, Wang X, Lv Q, Luo H, Zhang D (2022) Understanding wifi signal frequency features for position-independent gesture sensing. IEEE Trans Mob Comput 21(11):4156–4171. https://doi.org/10.1109/TMC.2021.3063135

Owfi A, Lin C, Guo L, Afghah F, Ashdown J, Turck K (2023) A meta-learning based generalizable indoor localization model using channel state information. arXiv e-prints, 2305–13453 https://doi.org/10.48550/arXiv.2305.13453

Pan M, Liu S, Liu P, Qi W, Huang Y, Zheng W, Wu Q, Gardill M (2023) In situ calibration of antenna arrays for positioning with 5g networks. IEEE Trans Microw Theory Tech 71(10):4600–4613. https://doi.org/10.1109/TMTT.2023.3256532

Ravi S, Larochelle H (2017) Optimization as a model for few-shot learning. In: International Conference on Learning Representations. https://openreview.net/forum?id=rJY0-Kcll

Shi Z, Cheng Q, Zhang JA, Da Yi XuR (2022) Environment-robust wifi-based human activity recognition using enhanced CSI and deep learning. IEEE Internet Things J 9(24):24643–24654. https://doi.org/10.1109/JIOT.2022.3192973

Snell J, Swersky K, Zemel R (2017) Prototypical networks for few-shot learning. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. NIPS'17, pp. 4080–4090. Curran Associates Inc., Red Hook, NY, USA. https://doi.org/10.3390/app11083329

Sung F, Yang Y, Zhang L, Xiang T, Torr PHS, Hospedales TM (2018) Learning to compare: Relation network for few-shot learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/CVPR.2018.00131

Tan S, Yang J, Chen Y (2022) Enabling fine-grained finger gesture recognition on commodity wifi devices. IEEE Trans Mob Comput 21(8):2789–2802. https://doi.org/10.1109/TMC.2020.3045635

Tzeng E, Hoffman J, Zhang N, Saenko K, Darrell T (2014) Deep Domain Confusion: Maximizing for Domain Invariance. arXiv e-prints, 1412–3474 https://doi.org/10.48550/arXiv.1412.3474arXiv:1412.3474 [cs.CV]

Venkatnarayan RH, Page G, Shahzad M (2018) Multi-user gesture recognition using wifi. In: Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services, pp. 401–413. https://doi.org/10.1145/3210240.3210335

Vinyals O, Blundell C, Lillicrap T, Kavukcuoglu K, Wierstra D (2016) Matching networks for one shot learning. In: Proceedings of the 30th International Conference on Neural Information Processing Systems. NIPS'16, pp. 3637–3645. Curran Associates Inc., Red Hook, NY, USA. https://doi.org/10.5555/3157382.3157504

Virmani A, Shahzad M (2017) Position and orientation agnostic gesture recognition using wifi. In: Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services, pp. 252–264. https://doi.org/10.1145/3081333.3081340

Wang Z, Guo B, Yu Z, Zhou X (2018) Wi-fi CSI-based behavior recognition: from signals and actions to activities. IEEE Commun Mag 56(5):109–115. https://doi.org/10.1109/MCOM.2018.1700144

Wang C, Liu J, Chen Y, Liu H, Wang Y (2018) Towards in-baggage suspicious object detection using commodity wifi. In: 2018 IEEE Conference on Communications and Network Security (CNS), pp. 1–9. https://doi.org/10.1109/CNS.2018.8433142

Wang F, Feng J, Zhao Y, Zhang X, Zhang S, Han J (2019) Joint activity recognition and indoor localization with wifi fingerprints. IEEE Access 7:80058–80068. https://doi.org/10.1109/ACCESS.2019.2923743

Wang Y, Yao Q, Kwok JT, Ni LM (2020) Generalizing from a few examples: a survey on few-shot learning. ACM Comput Surv (CSUR) 53(3):1–34. https://doi.org/10.1145/3386252

Wang Z, Huang Z, Zhang C, Dou W, Guo Y, Chen D (2021) CSI-based human sensing using model-based approaches: a survey. J Comput Des Eng 8(2):510–523. https://doi.org/10.1093/jcde/qwab003

Wang Y, Yao L, Wang Y, Zhang Y (2021) Robust CSI-based human activity recognition with augment few shot learning. IEEE Sens J 21(21):24297–24308. https://doi.org/10.1109/JSEN.2021.3111030

Wang D, Yang J, Cui W, Xie L, Sun S (2022) Airfi: Empowering wifi-based passive human gesture recognition to unseen environment via domain generalization. IEEE Transactions on Mobile Computing, pp. 1–12 https://doi.org/10.1109/TMC.2022.3230665

Wang D, Yang J, Cui W, Xie L, Sun S (2022) Caution: a robust wifi-based human authentication system via few-shot open-set recognition. IEEE Internet Things J 9(18):17323–17333. https://doi.org/10.1109/JIOT.2022.3156099

Wang Z, Ma M, Feng X, Li X, Liu F, Guo Y, Chen D (2022) Skeleton-based human pose recognition using channel state information: a survey. Sensors 22(22):8738. https://doi.org/10.3390/s22228738

Wang X, Ye W, Qi Z, Wang G, Wu J, Shan Y, Qie X, Wang H (2023) Task-aware dual-representation network for few-shot action recognition. IEEE Trans Circuits Syst Video Technol 33(10):5932–5946. https://doi.org/10.1109/TCSVT.2023.3262670

Wang Y, Wang Y, Liu Q, Zhang Y (2024) Dynamic wifi indoor positioning based on the multi-scale metric learning. Comput Commun 213:49–60. https://doi.org/10.1016/j.comcom.2023.10.022

Wei W, Yan J, Wu X, Wang C, Zhang G (2023) A meta-learning approach for device-free indoor localization. IEEE Commun Lett 27(3):846–850. https://doi.org/10.1109/LCOMM.2023.3241658

Wu D, Gao R, Zeng Y, Liu J, Wang L, Gu T, Zhang D (2020) Fingerdraw: sub-wavelength level finger motion tracking with wifi signals. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 4(1):1–27. https://doi.org/10.1145/3380981

Xiao R, Liu J, Han J, Ren K (2021) Onefi: One-shot recognition for unseen gesture via cots wifi. In: Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems. SenSys '21, pp. 206–219. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3485730.3485936

Xie Y, Wang H, Yu B, Zhang C (2020) Secure collaborative few-shot learning. Knowl-Based Syst 203:106157. https://doi.org/10.1016/j.knosys.2020.106157

Xue M, Chen Y, Gong X, Cao H, Zhang J, Zhang Q (2023) Meta-learning for human-centered wireless sensing: architecture, applications, and challenges. IEEE Network 37(1):88–94. https://doi.org/10.1109/MNET.114.2100774

Yang J, Zou H, Jiang H, Xie L (2018) Device-free occupant activity sensing using wifi-enabled IOT devices for smart homes. IEEE Internet Things J 5(5):3991–4002. https://doi.org/10.1109/JIOT.2018.2849655

Yang J, Zou H, Zhou Y, Xie L (2019) Learning gestures from wifi: a siamese recurrent convolutional architecture. IEEE Internet Things J 6(6):10763–10772. https://doi.org/10.1109/JIOT.2019.2941527

Yang C, Wang L, Wang X, Mao S (2022) Environment adaptive RFID-based 3d human pose tracking with a meta-learning approach. IEEE J Radio Freq Identif 6:413–425. https://doi.org/10.1109/JRFID.2022.3140256

Yang J, Zhou Y, Huang H, Zou H, Xie L (2022) Metafi: Device-free pose estimation via commodity wifi for metaverse avatar simulation. In: 2022 IEEE 8th World Forum on Internet of Things (WF-IoT), pp. 1–6. https://doi.org/10.1109/WF-IoT54382.2022.10152057

Yang J, Chen X, Zou H, Wang D, Xie L (2023) Autofi: toward automatic Wi-Fi human sensing via geometric self-supervised learning. IEEE Internet Things J 10(8):7416–7425. https://doi.org/10.1109/JIOT.2022.3228820

Yin G, Zhang J, Shen G, Chen Y (2022) Fewsense, towards a scalable and cross-domain wi-fi sensing system using few-shot learning. IEEE Trans Mobile Comput. https://doi.org/10.1109/TMC.2022.3221902

Yu C, Zhang D, Xie C, Lu Z, Hu Y, Li H, Sun Q, Chen Y (2022) Wifi-based human pose image generation. In: 2022 IEEE 24th International Workshop on Multimedia Signal Processing (MMSP), pp. 1–6. https://doi.org/10.1109/MMSP55362.2022.9949562

Zeng Y, Wu D, Xiong J, Zhang D (2021) Boosting wifi sensing performance via CSI ratio. IEEE Pervasive Comput 20(1):62–70. https://doi.org/10.1109/MPRV.2020.3041024

Zhang Y, Liu Q, Wang Y, Yu G (2022) CSI-based location-independent human activity recognition using feature fusion. IEEE Trans Instrum Meas 71:1–12. https://doi.org/10.1109/TIM.2022.3216419

Zhang Y, Zheng Y, Qian K, Zhang G, Liu Y, Wu C, Yang Z (2022) Widar3.0: zero-effort cross-domain gesture recognition with Wi-Fi. IEEE Trans Pattern Anal Mach Intell 44(11):8671–8688. https://doi.org/10.1109/TPAMI.2021.3105387

Zhang Y, Chen Y, Wang Y, Liu Q, Cheng A (2022) CSI-based human activity recognition with graph few-shot learning. IEEE Internet Things J 9(6):4139–4151. https://doi.org/10.1109/JIOT.2021.3103073

Zhang J, Chen Z, Luo C, Wei B, Kanhere SS, Li J (2022) Metaganfi: cross-domain unseen individual identification using wifi signals. Proc ACM Interact Mob Wearable Ubiquitous Technol 6(3):152–172. https://doi.org/10.1145/3550306

Zhang Y, Wang X, Wang Y, Chen H (2022) Human activity recognition across scenes and categories based on CSI. IEEE Trans Mob Comput 21(7):2411–2420. https://doi.org/10.1109/TMC.2020.3041756

Zhang X, Tang C, Yin K, Ni Q (2022) Wifi-based cross-domain gesture recognition via modified prototypical networks. IEEE Internet Things J 9(11):8584–8596. https://doi.org/10.1109/JIOT.2021.3114309

Zhang L, Wu S, Zhang T, Zhang Q (2023) Learning to locate: adaptive fingerprint-based localization with few-shot relation learning in dynamic indoor environments. IEEE Trans Wireless Commun 22(8):5253–5264. https://doi.org/10.1109/TWC.2022.3232858

Zheng L, Bi S, Wang S, Quan Z, Li X, Lin X, Wang H (2023) Resmon: domain-adaptive wireless respiration state monitoring via few-shot Bayesian deep learning. IEEE Internet Things J 10(23):20914–20927. https://doi.org/10.1109/JIOT.2023.3284407

Zhou Q, Yang Q, Xing J (2022) Enabling efficient wifi-based occupant behavior recognition using insufficient samples. Build Environ 212:108806. https://doi.org/10.1016/j.buildenv.2022.108806

Zhou Y, Zhu A, Xu C, Hu F, Li Y (2022) Perunet: deep signal channel attention in UNET for Wifi-based human pose estimation. IEEE Sens J 22(20):19750–19760. https://doi.org/10.1109/JSEN.2022.3204607